LUISS

Course of

SUPERVISOR

CO-SUPERVISOR

CANDIDATE

Academic Year

Contents

Abstract

The concept of data, its utilization and its evolution in the economy have sparked the interest of researchers for a long time and right now we are at a turning point for what we want and expect from this resource. The consequences of limited regulation are before everyone's eyes, they have produced massive companies that we are furiously trying to govern with late and sometimes outdated legislation, the rights to privacy and data protection systematically challenged and people not sufficiently informed on what companies do with their data. At the same time, the data sharing market still has untapped potential and this thesis provides a comprehensive exploration into its multifaceted landscape. Setting the stage with foundational terminologies such as data spaces, platforms, the FAIR principles, and the critical role of data intermediaries, the study creates an understanding of the data sector's bedrock. With this foundational knowledge, the thesis shifts the focus on the prevailing legislative frameworks that shape Europe's data sector.

The focal point of the analysis zeroes in on the Data Governance Act and the Data Act, highlighting challenges within the current data space model and underscoring trust and interoperability as the cornerstones for a successful data-sharing ecosystem. As a supplement to the academic discourse, insights from interviews with representatives from the International Data Spaces Association (IDSA) and Salus Coop, are interwoven to provide real-world perspectives and ground the analysis in practical contexts.

Concluding the study, it emerges that while the data-sharing sector shows promise, it remains encumbered by challenges, especially in areas of legal ambiguity within guiding legislation such as the Data Governance Act. The act, although envisioned to bolster the data-sharing arena, has been subject to scrutiny for potential ambiguities and uncertainties. This leads to a call for clarity and refinement to foster trust, innovation, and a robust data-sharing framework. Through this endeavor, the thesis serves as a beacon for understanding the contemporary state and potential trajectory of the data sharing sector in Europe, aspiring to contribute to further studies and shape the ongoing discourse in this dynamic field.

Introduction

In recent years, the data sharing sector has emerged as a pivotal area of interest in Europe, characterized by rapid developments and significant challenges. This thesis endeavors to delineate the intricate landscape of data sharing in Europe, adopting a an approach that encompasses an exploration of key terminologies, an overview of the existing legislative framework, and an in-depth analysis of these foundational acts overseen data exchanges.

Initially, the thesis elucidates essential terminology that forms the bedrock of understanding the data sector, including an in-depth exploration of concepts such as data spaces, platforms, the FAIR principles, and the role of data intermediaries. This foundational knowledge serves as a precursor to a comprehensive overview of the prevailing legislative frameworks governing the data sharing sector, providing a contextual background against which the subsequent analysis is framed.

In the third and final segment, the focus narrows to a critical examination of the Data Governance Act and the Data Act. This section delineates the inherent challenges posed by the current data space model, emphasizing the pivotal role of trust and interoperability in fostering a thriving data sharing ecosystem. Through an analytical lens, the study scrutinizes the implications of these acts, their potential to address existing problems, and the avenues they create for fostering a resilient and robust data sharing sector in Europe.

Supplementing the analytical discourse is a concise section embodying insights gleaned from interviews conducted with representatives from the International Data Spaces Association (IDSA) and SalusCoop. This section enriches the narrative by incorporating practical perspectives and expert insights, thus providing a rounded view of the current state of the data sharing sector in Europe.

Through a confluence of theoretical analysis and practical insights, this thesis aspires to contribute a nuanced and well-rounded perspective on the evolving data sharing sector in Europe. This research aims to contribute to the ongoing discourse in this domain, possibly aiding in fostering informed discussions and facilitating a deeper understanding of the sector. It hopes to offer insights that might be useful in shaping future studies, subtly guiding the sector towards a path of gradual innovation and increased trust and interoperability.

The terminology of data

2.1 Data

The term *data* refers to "any digital representation of acts, facts or information and any compilation of such acts, facts or information, including in the form of sound, visual or audiovisual recording".[1] The issue of analyzing the implication of data registering thoughout history, especially since the age of digitalization, has been discussed at length by the literature, and one aspect to analyze concerns the connection between data and information-as-thing[2]. In particular, the terms *data*, *data object*, *information*, and *knowledge*, are closely linked to one another.[3] Data, taken as is, without context and form, are mere facts and occupy the lowest step in the knowledge hierarchy for most of the literature.[4]

When they are structured, they become information and then knowledge, but Tuomi (1999) provides a different view on the issue stating the hierarchy should be reversed because "data can emerge only if a meaning structure, or semantics, is first fixed and then used to represent information". Hence data are created from information after it is dissected and categorized into a predefined structure. This is because "data is a set of discrete, objective facts about events [...] Data describes only a part of what happened; it provides no judgment or interpretation and no sustainable basis of action [...] Data says nothing about its own importance or relevance".[5]

Data can be defined as cultural records and human-made artifacts whose primary purpose is to store and transmit intangible information, such as knowledge. As a

[1] Regulation (EU) 2022/868 - Data Governance Act

[2] Ilkka Tuomi (1999). "Data is more than knowledge: Implications of the reversed knowledge hierarchy for knowledge management and organizational memory". In: *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. 1999. HICSS-32. Abstracts and CD-ROM of Full Papers*. IEEE, 12–pp.

[3] Cristina Alaimo and Jannis Kallinikos (2022). "Organizations decentered: Data objects, technology and knowledge". In: *Organization Science* 33.1, pp. 19–37.

[4] Ilkka Tuomi (1999). "Data is more than knowledge: Implications of the reversed knowledge hierarchy for knowledge management and organizational memory". In: *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. 1999. HICSS-32. Abstracts and CD-ROM of Full Papers*. IEEE, 12–pp.

[5] In Tuomi (1999) cites Davenport, Prusak, et al. (1998) to give a description of data that is shared by ost of the literature.

pervasive resource, data serves as a medium through which organizations come to know and act upon the contingencies they confront. They exist only as records and are an essential component of knowledge objects, forming a crucial part of the organizational decision-making process. Moreover, data are the means through which algorithms operate in the world, highlighting their growing importance in various fields and industries.

It is important to note that data are produced by a vast infrastructure of knowing and act as placeholders for organizational resources, making them visible and accessible in novel ways. As such, data can foster several novel knowledge and organizational processes, making them an indispensable tool for modern organizations. Therefore, "data are no longer a secondary component of administrative support but a pervasive *resource* and *medium* through which organizations come to know and act upon the contingencies they confront"[6]. The increasing importance of data underscores the need for organizations to invest in data management and data analysis tools, which can help extract valuable insights and improve organizational decision-making. As such, the strategic use of data can be a significant driver of organizational success, leading to improved performance and increased competitiveness in the marketplace.

When they are digitized, data objects represent the fundamental cognitive units that enable the execution of more comprehensive knowledge management operations. Their significance is paramount, as other higher-order knowledge processes would be unattainable without them.

However, the functions performed by data objects are heavily influenced by the technical requirements and dependencies that result from their integration within a broader technological data management framework. As technical components, data objects are subject to a range of technical prerequisites that must be met in order to operate efficiently within this framework. Therefore, it is imperative to recognize the critical role that data objects play in enabling effective knowledge management operations while also considering the technical aspects that underlie their use.

The main characteristics of data objects is that they are content-agnostic, non-neutral and homogenizing. The production of data does not take into consideration nor the context, nor the content of what is recorded, which can lead to vastly different outcomes in terms of consequences for its use. Despite this data is non-neutral because of methods and sources utilized during data collection can introduce biases, and data often mirrors the preferences or biases of the group involved in its creation or collection, hence it might not embody a diverse perspective, thereby forsaking neutrality. Data is also inherently susceptible to manipulation or selective usage. In

[6]Cristina Alaimo and Jannis Kallinikos (2022). "Organizations decentered: Data objects, technology and knowledge". In: *Organization Science* 33.1, pp. 19–37.

a similar vein, there is also the concept of algorithmic bias that emerges when data, already tinged with biases, is used to train algorithms, which then perpetuate and even amplify these preconceptions in their outcomes. Recognizing the non-neutrality of data is a critical step in fostering more accurate, responsible, and equitable data usage in today's data-centric era.

2.2 Data management and data spaces

The biggest industries and service providers in the world heavily rely on data, hence the question of data management has gained importance as the amount of data generated and stored never ceases to rise. There is a wide range of data management solutions that vary in administrative proximity, the distance between the data sources and the data administration, and semantic integration, how interrelated the data stored is and how well the definitions of each piece of data is coherent with one another. Data integration is the beginning of data processing, it is a mandatory step to make data valuable. The first problem of data integration is how to combine data from different sources and provide users with a single, unified view of the data. The architecture of data integration systems comprises of two components: the sources, the actual data, and the global schema, the association by which a coherent view of that data can be displayed.[7]. This means that the job of the DBMS is to create a global schema that can query all the sources at once in a language that maps the query language, which is the one of the global schema, to the various sources' elements, expressed in different languages ($\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$). To achieve the goal of relating the source and the schema, two main approaches have been developed: the first is called *global-as-view*, which directly links the two components expressing the schema; and a *local-as-view* system where the global schema is independent from the source and the link between these components are established by a definition of every source as a view. The source schema describes the structure of the sources, the global schema provides a view of such sources and the mapping is made by a set of assertion that links the two.

As shown in Figure 2.1, a query to $\mathcal{I}$ is posed through a language $\mathcal{L}_Q$ over an alphabet $\mathcal{A}_G$. It specifies what data to extract from the virtual data base. Virtual because is made up by different data sources put together by the integration system.

Database management systems (DBMS) are a generic repository for the storage and querying of structured data. They are "general-purpose software system that

[7]Maurizio Lenzerini (2002). "Data integration: A theoretical perspective". In: *Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pp. 233–246.

Table 2.1: Data integration framework

	Language	Alphabet
Global schema	$\mathcal{L}_\mathcal{G}$	$\mathcal{A}_\mathcal{G}$
Source schema	$\mathcal{L}_\mathcal{S}$	$\mathcal{A}_\mathcal{S}$
Query $q_\mathcal{S}$	$\mathcal{L}_{\mathcal{M},\mathcal{S}}$	$\mathcal{A}_\mathcal{S}$
Query $q_\mathcal{G}$	$\mathcal{L}_{\mathcal{M},\mathcal{G}}$	$\mathcal{A}_\mathcal{G}$

facilitates the processes of defining, constructing, manipulating, and sharing databases among various users and applications".[8] Nowadays it is difficult to find properly structured data. The challenges for managing heterogeneous data are the following:

- provide search and query capability;
- enforcing rules, integrity constraints, naming conventions, etc. . . ;
- tracking lineage;
- providing availability, recovery and access control;
- managing evolution of data and metadata.

There can be several DataSpace Support Platforms (DSSPs) serving the same data space.

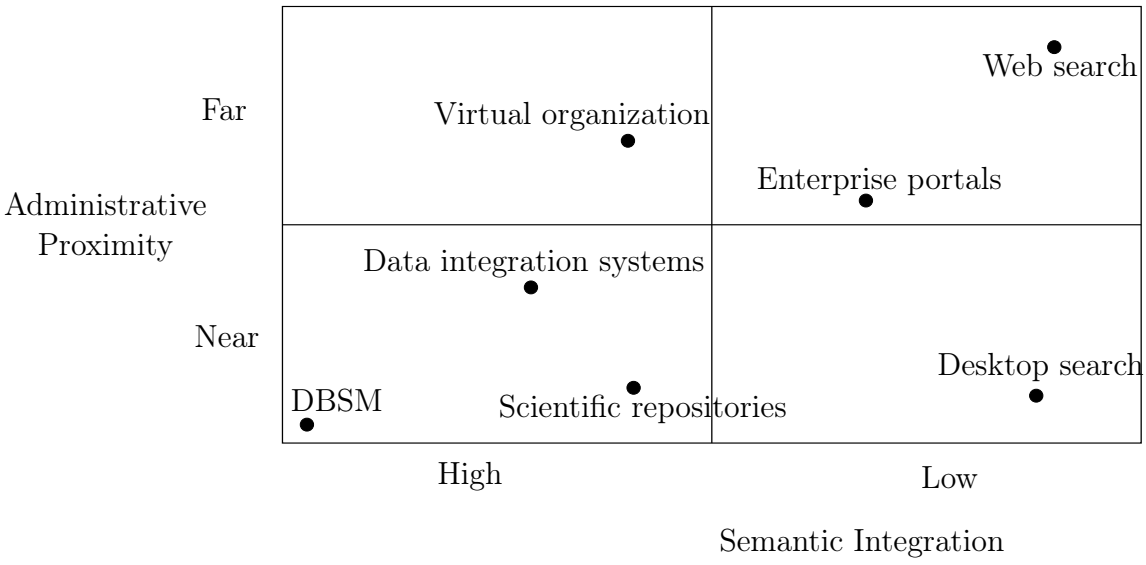

Figure 2.1: Data management solutions

For what concerns data management architecture, the categorization of data management solution comes via two dimensions: "administrative proximity" and

[8] Ramez Elmasri, Sham Navathe, et al. (2014). *Fundamentals of database systems*. Vol. 7. Pearson.

"system integration" (See Figure 2.1). Data spaces are a "data co-existence" approach in the sense that integration doesn't really matter in this context because the goal is to have basic functionality out of the gate.

At the core of the issue is the definition of what a data space is. The term was first used in Franklin, Halevy, and Maier (Dec. 2005) where the author claims a dataspace should "contain all of the information relevant to a particular organization regardless of its format and location, and model a right collection of relationships between data repositories. Hence, we model a data space as a set of *participants* and *relationships*". Participants are the individual data sources, such as XML repositories, relational databases, etc.... These sources can be stored or streamed, they can support expressive query languages, or limit the depth of search. Relationships between two or more participants can also be classified in the way they are constructed, i.e. we can defined a participant that is a copy of another, but structured differently; one created independently, and one that is the summa of sources A and B. Though the crucial difference between data spaces and data integration system is that the former does *not* require semantic integration to offer its services: data can be stored in different formats and still accessible by providing basic functions.[9] Another important note to this definition is that data can be integrated gradually, hence a dataspace can be refined and updated[10] as participants develop tools and standards get applied. In addition, DSSP is a perfect fit for collaboration between parties because the system is not in full control of its data, but the various participants are. Those features are crucial to give enterprises, small, medium and large, the time to adapt and adopt this kind of technology and instruments. This also opens up a business model of incremental plans to add functionality, consistency and durability to the DSSP,[11] much like modules of a digital platforms. In addition, DSSP are required to support all formats of data in the dataspace, while DBMSs aren't. The form of dataspace has been selected by the EU as the main tool to coordinate and create an Single Digital Market because it must offer the possibility of tighter integration of data, while in DBMSs this integration is immutable. All this characteristics are shown in Table 2.2.

Principles of dataspaces Dataspaces must enable accessing all the information of the desktop, whether they are implicit or explicit, and despite of the integration level, they will have to provide best-effort results. A dataspace-wide catalog is also

[9] Michael Franklin, Alon Halevy, and David Maier (Dec. 2005). "From Databases to Dataspaces: A New Abstraction for Information Management". In: *SIGMOD Rec.* 34.4, pp. 27–33.

[10] Yihan Wang, Shaoxu Song, and Lei Chen (Sept. 2016). "A Survey on Accessing Dataspaces". In: *SIGMOD Rec.* 45.2, pp. 33–44.

[11] Alon Halevy, Michael Franklin, and David Maier (2006). "Principles of Dataspace Systems". In: *Proceedings of the Twenty-Fifth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*. PODS '06. Chicago, IL, USA: Association for Computing Machinery, pp. 1–9.

Table 2.2: Comparison between DBMSs and DSSPs

Characteristic	DSSPs	DBSMs
data control	partial	full
modularity	yes	no
future-proof	yes	no
format support	full	partial
type of data	unstructured and semistructured	relational
query language	full-text search	structured query language (SQL)

necessary, as well as the support of data lineage. For the management of scientific data and collaboration, there is the requirement of creating collections and indexes over entities that span more than one significant source. Effective search mechanisms that can accept keyword queries and identify relevant structured sources capable of providing answers are essential in today's web searches. Additionally, the ability to combine answers from both structured and unstructured data sources has become increasingly important.[12]

There are two types of data spaces made for different purposes, the Industrial Data Spaces (IDS), and the Personal Data Spaces (PDS). The former are meant to provide a place where data can be share without worries of security and trustworthiness, geared primarily to the private sector and proprietary data, in short, a "trusted data sharing environment"[13]. The latter serves a personal information data management system where the user can control to which company share his/her personal data with.[14]

2.3 Two-sided markets and multi-sided platforms

Platforms, two- or multi-sided, are the core of *two-sided and multi-sided markets*. It is a class of businesses that creates value bringing together two or more market agents.[15]

In some markets, profit-seeking manufacturers who wish to remain competitive

[12] Alon Halevy, Michael Franklin, and David Maier (2006). "Principles of Dataspace Systems". In: *Proceedings of the Twenty-Fifth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*. PODS '06. Chicago, IL, USA: Association for Computing Machinery, pp. 1–9.

[13] Edward Curry, Simon Scerri, and Tuomo Tuikka (2022). *Data Spaces: Design, Deployment and Future Directions*.

[14] Tuukka Lehtiniemi (2017). "Personal data spaces: An intervention in surveillance capitalism?" In: *Surveillance & Society* 15.5, pp. 626–639.

[15] David S Evans and Richard Schmalensee (2013). *The antitrust analysis of multi-sided platform businesses*. Tech. rep. National Bureau of Economic Research.

view to get two separate groups of users 'on board'.[16]

Example. Case of console and videogame producers. No developer produces for a console that has no gamers and no gamer would buy a console with no game developed. The solution to this problem is to 'choose a price structure and not only a price level'[17].

In this example above, output can increase 'by charging more to one side and less to the other relative to what the markets delivers'[18]. This means that one side, called 'the money side' (gamers), will be called to cross-subsidize the participation of the 'subsidy-side' (developers) In this manner, the 'decomposition or allocation' of the total price between the two sides will affect output'.
There also must be *indirect network (cross-platform) externalities* or *effects*, i.e. users' participation on one side increases the participation of users on the other side and vice-versa. Also, the users must be prevented from negotiating away the platform's price.[19] Though this part of the "theorem" has been ignored in more recent scholarship.[20]

Two-sided market theory builds upon other IO concepts:

1. *network externalities* – the individual utility that one user derives from a good may raise with the number of other users that consume it. A prime example of this concept can be found in social media: the more users are active on the platforms, the higher the value of the platform.

2. *Coasean economics* – the Coasean theory tries to counterbalance the harmful effects on others caused by the action of firms. Theory was that **if** there is **no transaction cost**, then private bargaining can lead to an **optimal allocation** of resources.

For what concerns Coasean markets, Coase admits the possibility of alternatives when transactions costs are too high. "It is clear that Coase had foreseen [...]the role of platforms as a 'social arrangement' likely to resolve externalities"[21]. Also the two-sided market theory is not incompatible with Coasian bargaining: if we think of platforms as *social arrangements* that solve parties' inability to conclude bilateral

[16]See Jean-Charles Rochet and Jean Tirole (2003). "Platform competition in two-sided markets". In: *Journal of the european economic association* 1.4, pp. 990–1029 and Jean-Charles Rochet and Jean Tirole (2006). "Two-sided markets: a progress report". In: *The RAND journal of economics* 37.3, pp. 645–667

[18]Ibid.

[19]There must be transactions costs preventing 'the bilateral setting of prices between buyer and seller'

[20]Dirk Auer and Nicolas Petit (2015). "Two-sided markets and the challenge of turning economic theory into antitrust policy". In: *The Antitrust Bulletin* 60.4, pp. 426–461.

[21]Ibid.

transactions, then if users re-engineer the platform's pricing decisions these social arrangement can lead to Coasian bargains between users on each side of the platform. The following are the three most popular definitions of two-sided markets:

- market is two-sided if the platform can affect the volume of transactions by charging more to one side of the market and reducing the price paid by the other in an equal amount; in other words, the price structure matters, and platforms must design it so as to bring both sides on board[22]

- a multi-sided platform has "(a) two or more groups of customers; (b) who need each other in some way; (c) but who cannot capture the value from their mutual attraction on their own; and (d) rely on the catalyst to facilitate value creating interactions between them"[23]

- *some kind* of interdependence or externality between groups of agents that are served by an intermediary[24] – which is the most estensive definition in the literature[25]

The difference in those definitions has wide implications in what we can and cannot consider a two-sided market: in Rochet and Tirole (2003), two-sided markets are possible only where Coasian bargaining is impossible, but the other two definitions admit is. Following this definition, a supermarket can be classified as a two-sided market since there is no negotiation between suppliers and consumers, and the retail prices are determined solely by the supermarket. On the other hand, in the case of shopping malls, consumers and stores have the ability to influence pricing decisions by negotiating over the retail prices set by the stores. In contrast to this, the two definitions from Evans and Schmalensee (2013) and Rysman (2009) consider both, supermarkets and shopping malls, to be two-sided markets because they "solve a transactional problem between suppliers and consumers" and "address an indirect network externality". Following this example, some problems emerge: if we consider that it is usually not the supermarket that makes the price, but it is the supplier via distribution contracts, then we have a one-sided market under the Rochet and Tirole (2006) definition. For the other two definition, this is not a problem. In addition, beyond contractual restrictions, other factors like platform's governance structure, the legal system, etc... may have a decisive impact on whether to classify a market

[22] Jean-Charles Rochet and Jean Tirole (2003). "Platform competition in two-sided markets". In: *Journal of the european economic association* 1.4, pp. 990–1029.

[23] David S Evans and Richard Schmalensee (2013). *The antitrust analysis of multi-sided platform businesses*. Tech. rep. National Bureau of Economic Research.

[24] Marc Rysman (2009). "The economics of two-sided markets". In: *Journal of economic perspectives* 23.3, pp. 125–43.

[25] Dirk Auer and Nicolas Petit (2015). "Two-sided markets and the challenge of turning economic theory into antitrust policy". In: *The Antitrust Bulletin* 60.4, pp. 426–461.

as two-sided or not.

However, using definitions such as the last two mentioned above can lead to over-inclusiveness, resulting in errors when applied to antitrust policy. In addition, such definitions may fail to acknowledge the existence of a two-sided market where the cross-group externality is not immediately observable, thereby resulting in under-inclusiveness. To prevent under-inclusiveness in Rochet and Tirole's definition, it is essential to evaluate the Cosasian pass-through to determine whether the market is two-sided, but the threshold for doing so is not predetermined and up to the researcher.

Table 2.3: Comparison of two-sided market definitions. Source: Auer and Petit (2015)

	Rochet and Tirole (2006)	Evans and Schmalensee (2013)	Rysman (2009)
payment systems	Y	Y	Y
video game consoles	Y	Y	Y
operating systems	Y	Y	Y
online recruitment	N	Y	Y
airports	N	Y	Y
shopping malls	N	Y	Y
automobile engines	N	N	Y
highways	N	N	Y

Digital platforms. The term 'digital platform' refers to a concept built upon the definition given by Tiwana, Konsynski, and Bush (2010) of 'software-based' platforms as "extensible codebase of a software-based system that provides core functionality shared by the modules that interoperate with it and the inter-faces through which they interoperate". The slightly modified version incorporates the presence of 'modules', which extend the functionality of the software, and 'interfaces' that share that core functionality with the system.[26][27]

Table 2.4: Definitions of core concepts. Source: Tiwana, Konsynski, and Bush (2010)

Concept	Definition

[26] Ahmad Ghazawneh and Ola Henfridsson (2015). "A paradigmatic analysis of digital application marketplaces". In: *Journal of Information Technology* 30.3, pp. 198–208.

[27] For a recollection of digital platform research review see De Reuver, Sørensen, and Basole (2018)

Digital platforms	Software-based external platforms consisting of the extensible codebase of a software-based system that provides core functionality shared by the modules that interoperate with it and the interfaces through which they interoperate (Ghazawneh and Henfridsson (2015)).
Modules	An add-on software subsystem that connects to the platform and add functionality to the platform.
Interfaces	Specifications and design rules that describe how the platform and modules interact and exchange information.

2.4 Data Cooperatives

Data intermediaries are entities that institutions and scholars has tried to categorized in several ways[28]. It has been argued that those categorization can be irrelevant if we assume the broader definition of 'data intermediary' as an entity that 'enables the sharing of data between data holders and data users'.[29] Data cooperatives are a type of data intermediary that began operating on the markets in the 2010s. In Table 2.5 you can see some examples. Their creation was proposed by several experts, but the

Table 2.5: Data cooperatives typology

cooperative	motivation/purpose	data gathering and use
polypoly	personal data	no access to members data
Drivers' Seat	gig economy	direct work with data
MIDATA	health	N/A
Salus	health	no access to members data
SAOS	agriculture	direct work with data

economic model was not, and still isn't, able to spread and thrive. This is believe to be caused by several factors: lack of incentives, a market already saturated with for-profit organizations and the characteristic that, if data is nonrivalrous, i.e. the same data can be provided to multiple intermediaries, its economic value collapses.[30] Data cooperatives are characterized by three key features[31]:

[28]Heiko Richter (2023). "Looking at the Data Governance Act and Beyond: How to Better Integrate Data Intermediaries in the Market Order for Data Sharing". In: *GRUR International*, ikad014.

[29]Ibid.

[30]Shota Ichihashi (2021). "Competing data intermediaries". In: *The RAND Journal of Economics* 52.3, pp. 515–537.

[31]Thomas Hardjono and Alex Pentland (2019). *Data Cooperatives: Towards a Foundation for Decentralized Personal Data Management.*

- The members of the cooperative own and control their personal data. It is him/her who is in charge of the actions to take on them and the data is collected in a *personal data store*
- The cooperative have a fiduciary obligations to its members, thus must be administered and regulated by rules agreed by all members
- The objective of data cooperatives is to gain back control of the data and to seek appropriate remuneration for that data. Though this last goal is the lesser one, the most important step remains the ability of the data source (the physical person) to regain its ownership over personal data.

Their ecosystem, summarized in Figure 2.2 is constituted by (i) data cooperatives as legal entities (ii) individuals, the people actually owning the data (iii) third parties that interact with this data, referred to as *queriers*. To maintain and safeguard the ownership of data by members, contracts with third parties must include a prohibition from accessing or copying the data, which is especially advisable in cases where the cooperative outsources its IT services to third parties.[32] Data cooperatives offer a different way of looking at data sharing by private citizens, they empower the consumer and shift the balance of power in data exchanges. They are a valuable ally in fighting the current structure of the data market where the user is almost powerless when confronted with the enormous amount of data extraction it is forced to agree to.

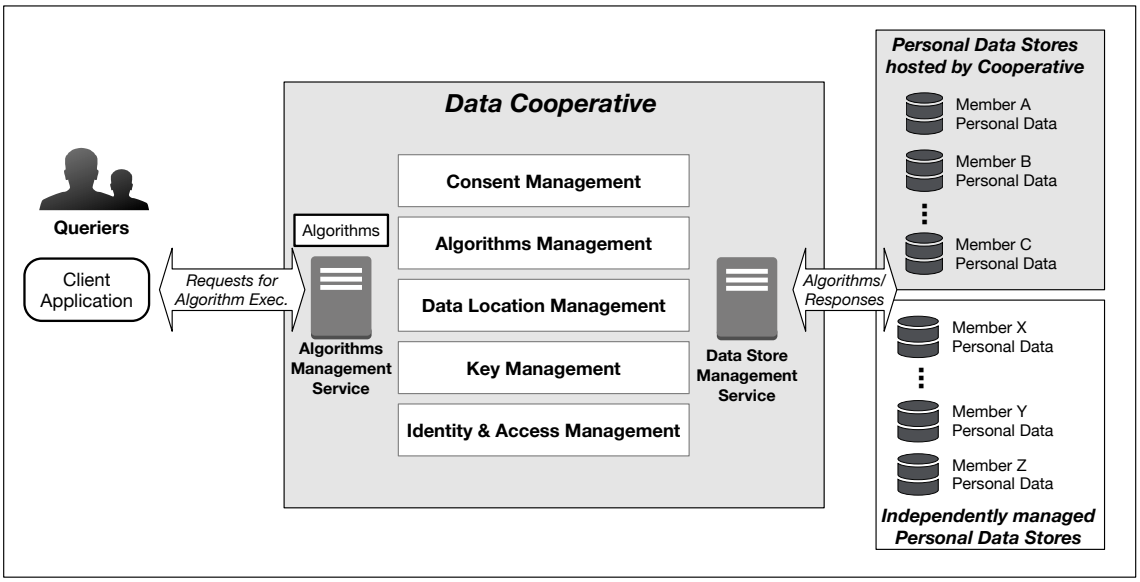

Figure 2.2: Overview of the Data Cooperative Ecosystem

[32]For example, SalusCoop has signed an agreement with a blockchain company in order to guarantee control to data access. See Section 4.4

EU approach to the data issue

3.1 Brief history of the EU approach to technology

The beginning of the digital market can be pinpointed by the Fifth Framework Programme, where the reference to the *information society*, a term always present in the programmes of the EU since 1995.

With the Delors Commission, there was a pursuit of research and development, new technologies, and what was called the information society.[1] It started after the eurosclerosis and the lagging competitiveness between Europe and USA and Japan that Delors decided to combat in the field of technological strength, hence one of the priorities of the commission was the development of new technologies. That terms was intended to be broad and was not limited to the information and communication technology[2] when it was introduced in the White Paper on the Completion of the Internal Market.[3] A passage that ties the current situation of the EU to the one in the mid-80s is when talking about the internal market, Delors pointed out that it would foster "cooperation between firms in the high technology sector to enable them to provide the technical strength [...] to compete with world-leading giants"[4]

The Single European Act (SEA) represented a pivotal point in this regards because it was explicit in the objective entitling *Research and Technological Development*

The obstacles in the path of completing the Digital Single Market are of three types: physical, technical and economic. There has been a widely insufficient infrastructure for internet access in less accessible areas, the so-called *white areas*, as well as geo-blocking[5], different levels of access depending on users' digital skills, the lack of high-speed Internet infrastructure at the European level and a geographical discrimination in e-commerce. Some of these issues are still present, but some have been mostly tackled.

[1] Citations in these sections are taken by speeches of Jaques Delors reported in Mirela Mărcuț (2017). *Crystalizing the EU digital policy: an exploration into the digital single market.* Springer

[2] The fields of application of these "new technologies" cited in the document where information and data processing services, computerized marketing and distribution services

[3] COM/85/0310

[4] Jaques Delors (1985). *The Dignity of Work and the Reasons for Peace.*

[5] the availability of a film on a digital platform, for example, can depend on the location of the user

The European Union faces challenges in tearing down technical barriers due to national approaches to standardization in the ICT sector, resulting in poor portability and interoperability. While Directive 2019/2161[6] aims to address consumer protection rights, uncertainties still remain regarding user rights within the DSM. Two issues the EU struggles with are inadequate security, which leads to low consumer trust, and the lack of European-level online platforms. The Data Governance Act aims to address the latter by creating Common Data Spaces. The EU has already addressed the issue of high roaming costs with the implementation of Regulation (EU) 2022/612.

The law regulating the digital market at the European level are not a cohesive framework, which undermines and hinders the development of a proper single markets, both on the demand and the supply side. The current European Commission is charging forward with its initiatives putting at the core of its action several pieces of legislation, some of which have already been approved[7], some others are still in debate[8].

Table 3.1: Issues and EU legislation on data.

Main issues in the data economy	**Legislation**
Lack of free flow and insufficient protection of personal data	GDPR
Lack of free flow of non-personal data/data localization requirements	FFoD Regulation
Lack of trust in data intermediaries	Data Governance Act
Insufficient availability of public sector data for re-use	Open Data Directive and DGA
Imbalances caused by the market power of gatekeepers	Digital Markets Act
Owners of connected products do not get value of their data	Data Act
Contractual imbalance between data holders and data users in data access and use that cannot be solved by competition law	Data Act
Insufficient means to access private sector data by public sector bodies in exceptional situations	Data Act

[6]This directive emended Directive 93/13/EEC and Directives 98/6/EC, 2005/29/EC and 2011/83/EU of the European Parliament and of the Council as regards the better enforcement and modernisation of Union consumer protection rules

[7]Digital Markets Act (Regulation (EU) 2022/1925), Digital Service Act (Regulation (EU) 2022/2065), Data Governance Act (Regulation (EU) 2022/868)

[8]Chips Act (COM/2022/46), Cybersecurity Act (Regulation (EU) 2019/881), Data Act (COM/2022/68)

Lack of interoperability between cloud services and hurdles to effective switching between providers across the market (beyond gatekeepers)	Data Act
Lack of interoperability	EIF and DGA (for the part of the EDIB)

In conclusion, EU's pursuit of a Single Digital Market has a long history, and despite the efforts put into it, there are still vast areas where the market is not yet fully developed, but the legislative work has been present and able to cover its portion of responsibility, as shown in Table 3.1.

3.2 Interoperability

The EU has been focusing on interoperability for more than two decades, the first initiative was launched in 1999 with Decision 1719/1999/EC[9], and the EU has since then been endorsing and funding initiatives that aimed at creating, fostering, and utilizing interoperability solutions throughout the EU. The main problem of data integration is the lack of standards and common practices about what to use in classifying data: data sources are often inconsistent and mutually

Another EU initiative to ensure the reuse of data by the public administration was the eGovernment Action Plan 2016-2020. It was developed with the aim of providing citizens across the EU with the full benefits of digital public services. To achieve this, the plan outlines several objectives, such as the establishment of a Digital Single Gateway that would enable users to access all necessary information, assistance, and problem-solving services across borders.

In addition, the plan seeks to interconnect business and insolvency registries and link them to the eJustice portal, which would function as a one-stop-shop for such services. The once-only principle for businesses across borders is also proposed, which would require them to provide information to a public administration in one EU country only and, if permitted, it would be reused in other countries.

Moreover, the eGovernment Action Plan proposes to aid EU Member States in developing cross-border eHealth services such as e-prescriptions, and accelerate the transition to e-procurement while implementing the once-only principle in public procurement.

[9]Decision of the European Parliament and of the Council of 12 July 1999 on a series of guidelines, including the identification of projects of common interest, for trans-European networks for the electronic interchange of data between administrations (IDA).

To achieve these objectives, the eGovernment Action Plan outlines 20 actions that address key policy priorities such as modernizing public administrations using digital enablers, enabling mobility of citizens and businesses through cross-border interoperability, and facilitating digital interaction between administrations and citizens or businesses for high-quality public services.

It is worth noting that additional actions may be proposed by various stakeholders, including the Commission, EU countries, and public administrations at all levels. Overall, the eGovernment Action Plan is a significant step towards improving the provision of digital public services throughout the EU.

3.2.1 FAIR guiding principles

The FAIR principles has been developed in the early 2010s, first appeared in the literature with Wilkinson et al. (2016), and subsequently developed by each sectors' scientific literature. To the point where some of the main proponent of these principles decided to publish a follow-up to further explain the meaning and implications of them to the public and the scientific community at large.[10]

FAIR principles

- Findability
  - F1. (meta)data are assigned a globally unique and persistent identifier
  - F2. data are described with rich metadata (defined by R1 below)
  - F3. metadata clearly and explicitly include the identifier of data it describes
  - F4. (meta)data are registered or indexed in a searchable resource
- Accessibility
  - A1. (meta)data are retrievable by their identifier using a standardized communication protocol
    - A1.1. the protocol is open, free, and universally implementable
    - A1.2. the protocol allows for an authentication and authorization procedure, where necessary
  - A2. metadata are accessible, even when the data are no longer available
- Interoperability
  - I1. meta(data) use a formal, accessible, shared, and broadly applicable language for knowledge representation
  - I2. meta(data) use vocabularies that follow FAIR principles
  - I3. meta(data) include qualified references to other meta(data)
- Reusability
  - R1. meta(data) are richly described with a plurality of accurate and relevant attributes

[10] Annika Jacobsen et al. (2020). *FAIR principles: interpretations and implementation considerations.*

R1.1. meta(data) are released with a clear and accessible data usage license
R1.2. meta(data) are associated with detailed provenance
R2. meta(data) meet domain-relevant community standards

The ultime objective of these principles is to have data that is perfectly machine readable and can thus be implemented by AI and machine learning. Hence, it should be easy for both humans and machines to retrieve and reuse data. In particular, "it should be possible for machines to merge the information into a richer, unified view".[11] The implication of applying such principles are wide and encompass most all field of life: for what concerns the principle of *findability*, the main issue is longevity of the identifiers, the definition of suitable metadata to ensure findability, and, the most complex one, the creation of a single repository of metadata for all fields. For *accessibilty*, the application of the principle implies that non-open access protocols need to be properly and fully documented, the definition of persistence policy for metadata. In Jacobsen et al. (2020), it is stressed that for *interoperability*, communities have to ensure that a data item present in multiple resources is not only the same, but every agent interprets it in precisely the same manner. The rationale of first point of the reusability principle (R1.) is to facilitate the comprehension and assessment, by both humans and machines, of the data's appropriateness for reuse within a specific task. It would require data provider to add more details in their metadata to contain also "operational instructions for reuse". It would require data provider to add more details in their metadata to contain also "operational instructions for reuse". Licensing that explains the use cases of the resource is crucial for individuals to utilize the resource legally, because its absence results in legal uncertainty that can either deter or prevent the reuse. There is currently no clear-cut definition in the relationship between data and metadata, and whether the licensing applies to the data, or to the metadata itself. Now it is up to the communities to make choices regarding the selection of usage licenses or licensing requirements for both reusable digital resources and their associated metadata. Additionally, they need to consider broader reuse possibilities that may extend beyond the initial intentions or expectations.

There is ultimately the need for communities to make choices and address implementation challenges to achieve optimal FAIRness in their respective domains. The role of convergence, guided by community-driven best practices and platforms, is crucial for achieving interoperability. Stakeholder communities should collaborate and share FAIR solutions to facilitate progress. Convergence requires agreement on the intentions of the FAIR principles and technological enablement through community-

[11] Annika Jacobsen et al. (2020). *FAIR principles: interpretations and implementation considerations.*

governed platforms. International coordination and a platform for convergence are necessary to guide this process effectively. The ultimate goal is to "establish an Internet of FAIR Data and Services"[12].

3.3 Data Protection and GDPR

The rights to privacy is paramount in our society, and it is two-sided, meaning it's both an opacity and a transparency right.[13]. As an *opacity right*, it aims at safeguarding the private sphere of individuals which cannot be transpassed, nor by other individuals, nor by the state or any other actor. This idea behind this reasoning is that it obligates others to abstain from interceding with the good, in this case privacy, that is protected. In addition, not only the State has to refrain from interference, but is has to enable the citizen to exercise its freedom. The transparency of this right entails that the processing of personal data must be done transparently and must be complying with the regulations for fair and lawful processing. Hence, the right to privacy is both a negative and a positive freedom.

In the realm of data protection, it is important to distinguish between the rights to privacy and data protection. While these two concepts overlap, they are distinct from one another. Specifically, the right to privacy also encompasses decisional privacy and bodily integrity, both of which do not involve the processing of personal data. On the other hand, the right to data protection pertains to the processing of personal information, such as banking details or personal addresses, for various purposes, unrelated to privacy.

Data protection in the EU is regulated by the General Data Protection Regulation[14] which is based on Article 16 of the TFEU where this right is enshrined. The objective of the regulation is twofold: assure the right to the protection of personal data and guaranteeing the free movement of them within the Union. In addition to this, the Police Data Protection Directive, the ePrivacy Directive as well as other pieces of legislation that tangentially touch the issue. Article 4 of the GDPR define processing as:

> any operation or set of operations which is performed on personal data, whether or not by automated means, such as collection, recording, organization, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction erasure or destruction.

[12] Annika Jacobsen et al. (2020). *FAIR principles: interpretations and implementation considerations.*

[13] Mireille Hildebrandt (2020). *Law for computer scientists and other folk.* Oxford university press.

[14] Regulation (EU) 2016/816

The notion of personal data contained in the GDPR is as broad as the processing one, involving "any information relating to an identified or identifiable person".[15] The issue is twofold: the first lies is the uncertainty of the term "identifiable", because the increase of availability and searchability of data can virtually render all of them personal. The criterion to establish whether a piece of data can be considered personal or not is the reasonable likelihood that a person can be singled out. This should be objective, since Recital 26 specifies that "all objective factors, such as the costs of and the amount of time required for identification" should be taken into consideration. Nevertheless, the case *Breyer v. Germany*[16], has shown that even the IP address, in some circumstances, can qualify as personal data. And the interpretation is not uniform across different data protection authorities. In similar cases concerning computer vision solutions to analyze costumers, Bavarian[17] and Dutch[18] DPA came to different conclusions because the former stated that since the time for processing data was so small, it did not have effect on de-anonymization. The latter took into consideration the content, the element and the purpose of the data processed and concluded it was in fact a case of personal data processing.[19]

Once define the scope of the regulation, the legal grounds for lawful processing of data are: (i) the data subject gave consent to the processing of his/her personal data for one or more *specific* purposes, and (ii) said processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except for whenever fundamental rights and freedoms are at stake. However, this point does not apply to public authorities in the performance of their tasks. I will not dive further into the details of the legislation, it suffices to say that the legitimate interest, as per Recital 47, is a controversial topic insofar it allows the controller to process personal data on the grounds of its economic interest. The principles of lawfulness, fairness and transparency are paramount to the architecture of data protection[20]. The purpose limitation refers to the fact that data collected and stored should be appropriate and relevant with regards to the purpose for which it has been

[15]The most used operative definition comes from the Guidelines of the Working Party, in opinion 04/2007 analyzing this definition. *personal data* is any information and statement about a person that either (a) "subjective" or "objective", i.e. referring to a particular person or is about that person, (b) it is likely to have an impact, even if small, on a person, (c) is used to influence the status or behaviour of a person, (d) and the processing of said data results in the identification of a person.

[16]CJEU, Case 582/14

[17]See in this respect Damian George, Kento Reutimann and Aurelia Tamò-Larrieux, 'GDPR Bypass by Design? Transient Processing of Data Under the GDPR' (2019) 9 International Data Privacy Law 285

[18]Dutch Data Protection Authority AP, Digital billboards standards framework, (2018)

[19]Laura Somaini (2020). "Regulating the dynamic concept of non-personal data in the EU: from ownership to portability". In: *Eur. Data Prot. L. Rev.* 6, p. 84.

[20]For a more in depth analysis see Franco Pizzetti (2016). *Privacy e il diritto europeo alla protezione dei dati personali: Dalla Direttiva 95/46 al nuovo Regolamente europeo*. G. Giappichelli Editore

stored, and the data subject has been adequately informed of it.[21]

This Regulation is complemented by Regulation (EU) 2016/680 which sets out the rules for public authorities on the processing of personal data in cases of criminal offenses. [22]

3.4 Open Data Directive

The first piece of legislation on minimum harmonization provision for the reuse of public sector information was introduced through the PSI Directive 2003/98/EC in an effort to foster a common legal framework for public sector information reuse within the European Union. The directive set minimum standards for the conditions under which public sector information *should* be made available for reuse, while allowing Member States to impose additional rules beyond these minimum standards. But this was only to promote this behaviour, there weren't binding rules.[23]

Subsequently, Directive 2013/37/EU further strengthened the legal obligations for the reuse of public sector information by mandating national entities to provide their open data to private stakeholders upon request. However, research data was excluded from the scope of these "open by default" and reuse obligations, which was justified on the grounds that research data was largely subject to intellectual property rights (IPRs).

This exclusion was subsequently altered by Directive 2019/1024, which extended the scope of the open data and reuse obligations to include research data[24] This legislative change reflects the growing recognition of the importance of research data as a valuable resource for scientific and economic advancement, and seeks to promote greater transparency and collaboration in this domain. However, Member States are still allowed to introduce exemptions to the reuse obligation following the rule "as open as possible, as closed as necessary". Public interest exceptions should be construed and interpreted narrowly to ensure the effectiveness of open access policies.

It is important to note that this Directive applies to *publicly funded* research, be it in the form of National open access policies must address research-performing and

[21] On this issue and on a proposed solution to verify the compliance see Basin, Debois, and T. Hildebrandt (2018).

[22] Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offenses or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA

[23] Sara Gobbato (2020). "Open Science and the reuse of publicly funded research data in the new Directive (EU) 2019/1024". In: *Journal of Ethics and Legal Technologies* 2.2.

[24] Research data are documents in digital form, aside from scientific publication, which are collected and produces in the course of scientific research activities, i.e. "statistics, results of experiments, survey results", etc.

research-funding organizations, including hybrid organizations. Hybrid organizations must comply with the Directive only for their publicly funded research activities and related research data under a functional approach. Data should be reusable if falling under these categories[25]:

- it is not subject to IPRs or third parties rights (legitimate commercial interest, knowledge transfer);
- they are publicly funded;
- researchers, and their institutions, have open sourced them through "institutional or subject-based repository";

Rec. 28 states that is up to member states whether to broaden the scope of action of the Directive to "data made publicly available through other data infrastructures than repositories". Hence, the crucial factor for the emergence of legal obligations regarding the reuse of research data under the Directive seems to be whether the data has been made available in a public repository or an open access journal, instead of a closed-access one. The characteristics of a high-value dataset are specified in Article 14, par. 2, indicating that the dataset must have the potential to produce a significant impact on the economy and society, serve the interests of a large number of people, facilitate revenue generation, and be capable of being integrated with other datasets. Annex I provides a list of thematic categories[26] to identify which datasets can be defined as having high-value, and if a datasets falls into one of the categories,[27] must be accessible to the public free of charge, be machine readable, provided via APIs and available as bulk download where relevant.

3.5 Regulation on the free flow of non-personal data.

Continuing the work on the Digital Single Market, the Junker Commission achieved the coming into law of Regulation (EU) 2018/1807 on a framework for the free flow of non-personal data in the European Union. Data processing is a core part of the data value chain, and the EU has already began to regulate some of its parts, as I mentioned in Section 3.1, effectiveness and efficiency are hampered by "data localization requirements put in place by Member States' authorities and vendor lock-in practices in the private sector".[28] The Regulation intends correct the legal

[25]Article 10, par. 2
[26]Commission Implementing Regulation (EU) 2023/138
[27]That can be adjusted to follow technological and market development
[28]Recital 2

uncertainty surrounding localization requirement sets out by national legislation, and any other requirement having the same effect. At the same time, the objective is to regulate restrictions made by the private sectors in the contracts, which can have the form of legal, contractual of technical ties/issues. This Regulation does not affect the legal framework of the protection of personal data, and, following the same principles of said Regulation, it forbids Member State's to impose restrictions or prohibition to the free flow of data, aside from when public security is at risk.[29] In order to avoid excessive burden on Member States and on the private sector, detailed rules are not established, following an approach of self-regulation for both parties.

The legislator emphasizes that the term data processes has to be intended in its broadest sense, hence including all intensities of processing: data storage, platforms and applications.[30]

For the implementation and the monitoring of compliance by Member States, in the transitional two years period, they should review their current legislation on the topic and communicate to the Commission which pieces of legislation they deem to be in compliance with the Regulation, as well as justification maintaining it in force.[31] In addition, a single information point should be set up where users and service providers can check the current requirements.[32]

On the topic of data porting for cloud services, this is another area left to self-regulation, "encouraged, facilitated and monitored by the Commission, in the form of Union codes of conduct which might include model contractual terms and condition".[33] Article 6 states these codes will be based on transparency and interoperability principles, and will take into account open standards. The codes will address various aspects, such as best practices for facilitating the switching of service providers and porting of data, minimum information requirements for professional users, certification schemes, and communication roadmaps. The Commission will work closely with all relevant stakeholders, including SMEs, start-ups, users, and cloud service providers, to ensure that the codes of conduct are developed collaboratively. The other tasks of the Commission are to submit a report of the implementation of this Regulation and publish guidance on how to handle datasets composed by both personal and non-personal data.[34]

As seen in Section 3.3, since the notion of personal and non-personal data is still blurry and varies across Member States, the very objective of creating a uniform data

[29]Recital 10

[30]Respectively Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS), Software-as-a-Service (SaaS).

[31]Article 4, par. 3.

[32]Alternatively, a central information point at the Union level should be set up with a different act.

[33]Recital 30

[34]Article 8, par. 3.

trading environment in the EU becomes difficult to achieve. As noted in numerous EU official documents, the risk of fragmentation is one of the major hurdles in the creation of a Digital Single Market.

Analysis of Data Act and Data Governance Act

4.1 Issues hindering the development of data sharing

In the rapidly evolving data sharing sector in Europe, several critical issues have come to the surface, necessitating urgent and thoughtful intervention. One of the most pressing concerns is the abuse of monopolistic market positions. This is a nuanced issue, as the repercussions of monopolistic control can vary substantially depending on whether the data in question serves as a complementary or substitutive element in secondary services. It has already began to being tackled by the Digital Markets Act and Digital Service Act, but strictly speaking in terms of data sharing, the DGA and DA are crucial pieces to this puzzle.

Simultaneously, the sector is grappling with substantial coordination problems, particularly where the reuse of data is concerned.[1] When a single or monopolistic entity controls complementary inputs required for data reuse, it generates a significant imbalance in bargaining power between the service providers and the consumers, hindering broader data sharing initiatives. An equitable solution could be fostered through the promotion of open market availability of these inputs, balancing power dynamics and enhancing collaborative efforts in the data sharing realm.

Adding to the complexity of the landscape are the uncontrolled externalities that have proven difficult to govern. Instances of data spillages involving sensitive or personal information, coupled with the inability to enforce bilateral data contracts with third parties where data ownership rights are ambiguous, are common. These challenges necessitate the implementation of stringent controls and regulatory frameworks that not only curb unauthorized data dissemination but also clearly delineate data ownership rights, fostering a safer and more controlled data sharing

[1] European Commission (2022). *Commission Staff Working Document. Impact assessment report accompanying the document Proposal for a Regulation of the European Parliament and of the Council on harmonized rules on fair access to and use of data (Data Act).*

environment.

Moreover, the sector is marred by pronounced information asymmetry, a phenomenon that has the potential to significantly distort the market. This imbalance often leads to a decrease in data protection, paving the way for exploitative practices. To mitigate these issues, it is imperative to foster transparency and ensure that all market participants have access to equal and unbiased information. Such initiatives could usher in a more secure and fair data market, encouraging growth and innovation while safeguarding individual rights and privacy.

For these reasons there is a necessity to provide market-based solutions that can tackle the oligopoly of large firms in this sector. Those solutions have to be respecting of the privacy and assure the parties that their data is safe and the data transaction they make happens in a secure environment. Market actors such as data trusts, industrial data platforms and data spaces can become the solutions for data sharing. These platforms offer the opportunity to reduce transaction costs by helping to standardize and make interoperable the data they trade. However, this is not a guaranteed outcome and platforms can derail in lock-in practices that are hard to dismantle once in place. [2] In addition, intermediaries should be neutral in their position, which is often not feasible because of the inherent bargain position advantage they have, since the platform is the primary method of connecting multiple data providers to data data users.

To help the market behave correctly from the beginning, a public intervention in setting the rules of the game could be beneficial, and if said light intervention does not produce the excepted results, then a more robust intervention is needed. Especially in the form of laws on data portability and accessibility.

Ex-post interventions are used in the case of sporadic market failure, and occur in two scenarios: when there is a refusal of access to data framed as an abuse of dominant position; or when a firm prices unfairly its data, by having a policy of self-preferential access to data. In the former case, there are some thresholds to consider the refusal a breach of Art. 102 TFEU, such as the indispensability of the data to provide a service, meaning that that sold data is essential and there are no close substitutes of the data.

These are exactly the topic where both the Data Governance Act and Data Act will intervene.

[2] We see now the advent of the Digital Market Act and Digital Service Act as an ex-post solution of dubious impact, but it is still a step in the right direction

4.2 Data Governance Act

4.2.1 Data Governance Act overview

The Data Governance Act, Regulation (EU) 2022/868, has to objective of facilitating data sharing and data altruism, creating a single market for data. Motivation to share data is limited because there is an interest of companies, who holds the data, to maintain advantage over its competitors, and there is a plethora of problems that concerns the law, the infrastructure and the public.[3] As an example, there might be breaches of intellectual property rights or data protection laws that could prevent more risk averse players from joining the market.[4]

Data in scope of the Regulation are the one not covered by the Open Data Directive, which include commercial and confidential data, personal data, and, more in general, *protected data*. Outside this scope are data held by public undertakings and hybrid public and private organization, but for the latter, only the data not related to the general interest are to be excluded from the scope of the regulation.[5]

The Regulation is without prejudice not only of the GDPR but of other sector specific Union Law relating to intellectual property rights (IPRs), security and law enforcement. The text takes also a forward-looking approach by anticipating the *compliance* with law on sectorial data spaces in at the EU level.[6].

Following this approach, there is a plethora of safeguards put in place to ensure the data is properly handled and utilized. While there are some explicit prohibition and rules for data processing in third countries and remote processing, the specific technical and legal requirements are not present. The letter of this legislation only provides that reuse in third countries is possible only when there is *contractual commitment* to the protection of data and the re-user accepts the jurisdiction of the Member State whose data came from in case of legal issues.[7]. The Commission will also lay down in delegated acts stricter conditions for the use of non-personal data included in specific categories[8] and, if a sizable number of requests concerning the reuse of data by a third country is presented to the Commission, it will have to adopt implementing acts that ensure the trustworthiness of said country via an examination procedure.[9] In addition, Artt. 27 and 28 guarantee to data subjects the rights to lodge a complaint to the relevant competent authority and subsequently

[3] See Subsection 4.2.3 for more details.

[4] Alina Wernick (2020). "Defining Data Intermediaries: A Clearer View Through the Lens of Intellectual Property Governance". In: *Technology and Regulation* 2020, pp. 65–77.

[5] There is explicit reference to research centers and research organization in Rec. 12

[6] Art. 1

[7] Art. 5(10)

[8] Rec. 24, Art. 6(13)

[9] Art. 5(12)

the rights to an effective judicial remedy before the court of the Member State where the data intermediation service provider or data altruism organization is located.

Data intermediation services involve facilitating the handling and processing of data between various parties. When a company offers these services within the European Union (EU), it must ensure legal representation in the Union if there is an evident intention to operate within its boundaries. It is essential to demonstrate a clear intent rather than merely making services accessible to ascertain compliance. Additionally, data intermediation service providers are required to follow a notification procedure to inform relevant authorities about their activities. To monitor adherence to regulations, a competent authority should be designated to oversee and ensure compliance with data governance and privacy laws in the EU. These measures aim to foster transparency, accountability, and lawful operations within the data intermediation sector while safeguarding data privacy and protection rights.

The competent authorities set up or designated by this Regulation should not affect the supervisory powers and competences of data protection authorities.[10] And here lies a potential problem of conflicting competences: how should be decided and adjudicated whether it's one authority of the other to have competence if the Regulation is unclear on these rules, the Commission should,and probably will, intervene with an Implementing act. Additionally, since this Regulation does not "create an obligation to allow the re-use of data held by public sector bodies",[11] but leaves the door open for Member States to legislate on the matter, no real progress is made from the Open Data Directive from an *open data* perspective.

In order to preserve the market and avoid the insurgence of monopolies, and anti-competitive behavior in general, Art. 4 prevent the signing of exclusive agreements for reasons outside the provision of products made in the general interest. This exclusive agreement has to follow the principles of transparency[12], equal treatment and non-discrimination as per national and Union law. Furthermore, the exclusive right's duration shall not surpass 12 months and any such contract in place before the 23rd of June 2022 will cease to be valid in any case by the 24th of December 2024.

Furthermore, a register of all recognized organizations with information on the entity's name, status, website, contacts and objectives should be available to the public and the Commission shall be notified of every addition in order to update the Union's register.[13]

Along with these registers, the Union wants to improve trust and boost data altruism assuring data subjects that their data will be treated fairly and transparently

[10]Recital 4
[11]Rec. 11
[12]Art.4 (5). The agreement shall be publicly available online.
[13]Art. 19(8)

setting 'specific requirements to safeguard rights and interests of data subjects and data holders with regard to their data' and aided by a consent form and a rulebook standard across the EU.

The former will be established via implementing act with the assistance of the European Data Innovation Board and all relevant stakeholders; its main feature will be the modularity to allow specific customization for each sector of the industry.[14]. The latter will supplement this Regulation by diving into details on (a) the 'appropriate information requirements' on how will data be used, the tools to revoke permission of such use and more general measures to prevent the misuse of the data shared; (b) the technical and security requirements; (c) a 'communication roadmap to raise awareness' on the topic of data altruism;[15] (d) relevant interoperability standards.[16]

European Data Innovation Board. The creation of the EDIB[17] will be the most significant bodies to intervene in this matter and will have a crucial role in determining the success of the endeavor towards an ever increasing use of data. It will be composed by the competent authorities for data intermediation services and data altruism organizations, among other EU bodies and will be divided into three subgroups: one specialized on data intermediation services and data altruism organization, the second will tackle standardization and portability, and the third will focus on data spaces. This partitioning of the Board follows the tasks laid out in Art. 30. I grouped them into 4 categories:

1. protection of data (e) and, more specifically, sensitive non-personal data (d), (m)

2. general standardization of practices for public sector bodies and competent authorities regarding re-use (a), for data altruism organizations (l)(b) and data intermediation services (c)(b), and interoperability more broadly (g)

3. data spaces (h) (f)

4. cooperation between Member States (i)(j)

4.2.2 Data intermediation services, or data sharing platforms

Art. 1(11) of the DGA defines data intermediation services as aimed at 'establishing a *commercial relationship*[18] for the purpose of data sharing, including for the purpose

[14]Art. 25

[15]This seems to be one of the most important topics that actors are concern about. See Section 4.4

[16]Art. 22

[17]Art. 29

[18]An emphasis is placed on commercial relationship since it is the legislator itself that points out in several instances this characteristic.

of exercising the rights of data subjects in relation to personal data', though excluding data processing services, intermediaries of copyrighted material, and services that are solely utilized by a single data holder to facilitate the use of their held data, or those employed by several legal entities within a restricted group.[19]

Data intermediation services play a pivotal role in promoting voluntary data sharing and facilitating non-discriminatory access to the data economy. They enable bilateral and multilateral data sharing through the establishment of platforms and databases for collaborative data usage. Moreover, these services contribute to the creation of infrastructure for seamless interconnection. Two critical principles in this context are data neutrality and the necessity of structural separation between data intermediation services and any other services, as outlined in Rec. 27 and 33.

For the purpose of this Regulation, data intermediation services are not: (a) services that obtain data from data holders and aggregate and enrich the data for the purpose of adding substantial value (b) services that intermediate copyright-protected material (c) dis that do not aim to establish commercial relationships

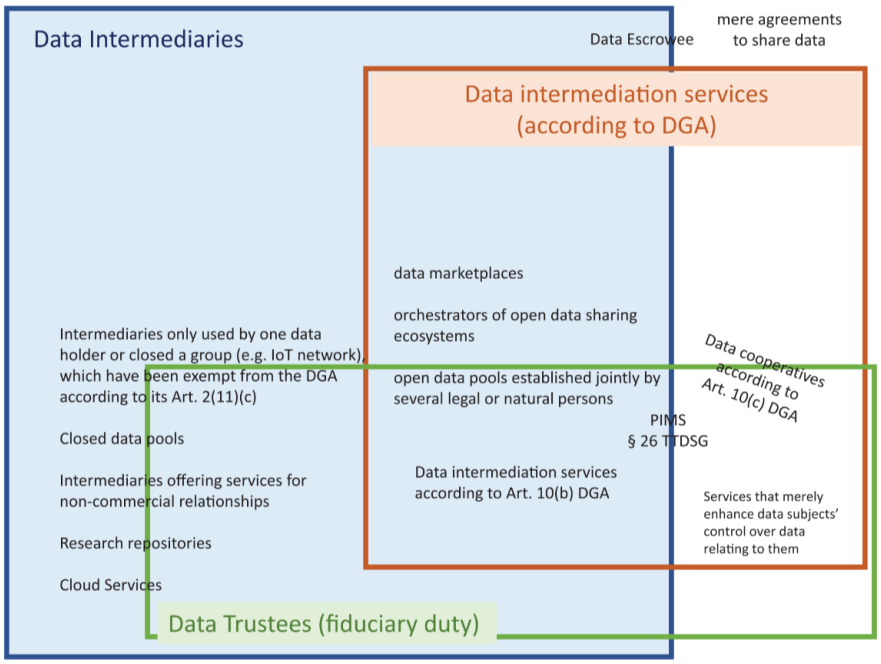

Figure 4.1: Data intermediation services differentiation. Source: Richter and Slowinski (2019)

Data intermediaries vary in: i openness; ii type of data (personal or non-personal); iii commercial or non-commercial; iv voluntary or mandated data sharing; v remunarates or does not remunerate the data holder; vi own interest or particular

[19]More specifically, the Regulation states 'including supplier or customer relationships or collaborations established by contract, in particular those that have as a main objective to ensure the functionalities of objects and devices connected to the IoT' since this will be taken care of by the Data Act.

duty, for instance, fiduciary duty for data trustees.[20]

A specific type of data intermediary is the data cooperatives whose main objective is to support its members in the exercise of their rights with respect to certain data and make informed choices before they choices and negotiate terms and conditions for data processing on behalf of their members.

Table 4.1: Data intermediation services

Service	Data Intermediation (Y/N)
Intermediation of copyright-protected content	N
Data spaces	Y
Public sector bodies in the data altruism organization	N
Cloud storage	N
Analytics	N
Data sharing software	N
Web browsers	N
Browser plug-ins	N
Email services	N

Despite this all intermediaries share the same principles as platforms. Their primary function is to match supply and demand, they need to attract participant who are willing to share their data, and participants who are willing to buy said data. Here lies also the first issue, how can the buyer assess the quality of data without looking at it directly, it need a third party, the platform, to do the vetting and rebalance the information asymmetry. To achieve this and sustain the platform, trust is the unavoidable requirement[21] because the stakes in data sharing for enterprises are extremely high, take for instance the case of trade secrets: what would happen if such data was to be accessed without permission? The consequence for the business would be potentially catastrophic. This Regulation tries to ensure that potential participants share their data in a safe environment and their rights are protected.

Furthermore, outside a vetting role that could include, for instance, overseeing transactions directly, or implement usage restrictions, there are technical tools that could strengthen trust, such as certification mechanisms, differentiated security levels, use of blockchain technology[22] and digital watermarking. Ultimately, trust is the most significant variable in the space.

[20]Heiko Richter and Peter R Slowinski (2019). "The data sharing economy: on the emergence of new intermediaries". In: *IIC-International Review of Intellectual Property and Competition Law* 50, pp. 4–29.

[21]European Commission and Everis (2018). *Study on data sharing between companies in Europe.*

[22]This instrument is in use by SalusCoop, see Section 4.4

4.2.3 Data Spaces

Data spaces are given a role of prominence in this regulation and it is clear that the EU is putting a lot of emphasis and effort in pushing this model. Data spaces are seen as one of the pillar of this new boost to the data sharing economy; they are the focus of three main objectives of the EDIB: (h)(g) to combat market fragmentation and (f) for what concerns sectoral and cross-sectoral standards for data spaces. Despite this, there are some issues surrounding data spaces that have to be addressed. In the following paragraphs, I will provide a brief overview and the solutions that are being proposed by the literature.

Firstly, the existing data lifecycle management models are not structured with data sharing as a central focus, thus the need to emphasizes the integration of services such as data cleaning and aggregation. These services are paramount for laying the foundation of a viable data economy, which is further complicated by the variety and intricacies associated with the different kinds of data available for sharing, requiring enhanced interoperability solutions. Secondly, the enforcement of data usage rights is critical to ensure digital sovereignty, helping data producers retain control over the their data, which necessitates the exploration of different ownership models and data rights management frameworks.[23]

Thirdly, there is a need for decentralized data sharing and processing architectures to have a scalable infrastructure ready to meet the future demands and volume of the market. For this reason it is also imperative to develop standard data exchange protocols that can foster such decentralized infrastructures. Simultaneously, addressing the challenges of data veracity, i.e. the accuracy, consistency, quality and trustworthiness of a data set, by incorporating traceable information about data origins and operations, thus increasing trust. The journey towards a resilient data-sharing ecosystem is not complete without addressing security concerns, emphasizing secure data access, confidentiality, and standardized security solutions across all nodes and participants in the data space, thus fostering a trusted network for exchanging and sharing closed data. Lastly, the industry still demands, rightly so, a meticulous focus on privacy protection, hence, encouraging continuous advancements in privacy-preserving technologies and exploring more flexible avenues for the integration of compliance solutions is a key endeavor.[24]

[23]Simon Scerri et al. (2022). "Common European Data Spaces: Challenges and Opportunities". In: *Data Spaces: Design, Deployment and Future Directions*, pp. 337–357.

[24]Ibid.

4.2.4 Conclusions

The use of delegated acts in this regulation is problematic because it touches issues that would have needed to be clear from the beginning. The Commission will legislate on two issues in particular: first, the rulebook to comply to for data altruism organizations in order to be officially registered and recognized in the Union, second, special conditions for the transfer of data that will be, *in the future*, considered highly sensitive. Further legislation currently missing and hinted in this Regulation revolves around data spaces, but at the time of writing, there is no official text.

4.3 Data Act

4.3.1 Data Act overview

The Data Act is set to be approved in the following weeks after fruitful trilogue negotiations that reached an agreement between the parts. For the purpose of this thesis, only Chapters II (Artt. 3-7), III (Artt. 8-12), IV (Artt. 13) and VIII are relevant.[25]

Chapter II - Business to consumer and business to business data sharing
Art. 3(2a) lists the information service providers are required to give the user. Out of this list, two pieces of information stand out: the 'nature, estimated volume and *collection frequency* of product data that the prospective data holder is expected to obtain'; and the 'nature and estimated volume of related services'[26] Art. 4(6) on the prohibition of use data to determine 'economic situation, asset and production methods' is wishful thinking at best, and blatant surrender to pressure at worst. And 6a talking about data holder not making available non-personal data to third parties for both commercial and non-commercial purposes is again wishful thinking. As happened to the legitimate interest in the GDPR, these provisions will be circumvented and will create the paradox where the consumer has rights that are systematically, but lawfully infringed upon.

On the right of the user to share data with third parties, Art. 5 tackles the overwhelming powers of gatekeeper, defined in the DMA, by forbidding to be considered as viable third parties to whom share data. It also extends to thrid parties the protection accorded to users against the use of data or readily available data to to unearth insights into various dimensions of the third party's commercial operations - be it economic status, assets, production methodologies, or usage patterns. This

[25] Chapter V focuses on business-to-government data sharing in situations of public emergency.
[26] Related services, as per Art.1(3), are digital services that are included in the product at the time of purchase and whose absence will lock out the user from one or more functions

prohibition is grounded in the potential risk that such data usage could potentially compromise the commercial standing of the third party in their active markets. At the same time, third parties shall not use the data received to *profile* the user, manipulate its judgment or make such data available for other third parties. To avoid anticompetitive behaviour, third parties are forbidden to use the data received to develop products that competes with the data holders'.

Since the EU has always been very conscious of the industrial economic structure in the Union and wants to create a competitive market for data, it provides some leeway to micro and small enterprises allowing them to disregard all aforementioned obligations.[27]

Chapter III - Obligations for data holders legally obliged to make data available The obligations stated in the chapter are applied where Art. 5 comes into force to make data available, and the non-compliance with such obligations deems null any contractual term. Obligation of the data holder (Art. 8):

- make data available to a data recipient
- holder shall agree with recipient terms for making that data available
- principle of non-discrimination between 'comparable categories of data recipient'
- prohibition of exclusive agreements
- making data available preserving trade secrets

Compensations for making data available (Art. 9):

- for micro, small and medium enterprises,compensations shall not exceed the cost of making the data available
- mandates clear communication and justification of the compensation asked by the data holder to the data recipient. It emphasizes a check and balance mechanism to ensure adherence to the mentioned rules and regulations

Technical protection measures and provisions on unauthorized use or disclosure of data (Art. 11): they should not be used by the data holder to hinder the user's right to effectively provide data to third parties. If a data recipient acted in bad faith to acquire data and used it for unauthorized purposes, it shall destroy the data and end its distribution.

Art. 8 details the duties that data holders should adhere to, fostering an atmosphere that facilitates easy data sharing while maintaining transparency and integrity.

[27] Art. 7

Under Art. 8, data holders are tasked with making valuable data available to recipients, thus promoting cooperative dynamics within the data sphere. Additionally, the data holder and recipient are required to come to a mutual agreement regarding the terms of data sharing, to foster a regulated and structured process. This directive is grounded in a commitment to non-discrimination, ensuring that all comparable categories of data recipients have a fair chance in the marketplace.

Moreover, the Act prohibits the formation of exclusive agreements, in an effort to prevent monopolistic tendencies and encourage competition. Concurrently, safeguarding trade secrets remains a critical duty of data holders, thus protecting intellectual assets and fostering trust in data transactions.[28]

Shifting focus to Article 9, the Regulation addresses the financial aspects of data sharing, particularly spotlighting smaller entities and non-profit research organizations as highlighted in the Provisional Agreement.[29]

On compensations, the legislation mandates that it should not surpass the actual costs incurred in making the data available. This approach aims to encourage affordability and inclusivity in data transactions. Furthermore, clear communication between the parties involved is emphasized, fostering a culture where compensation expectations are rational and justifiable. Proceeding to Article 11, it introduces stringent rules regarding technical protection measures and delineates the consequences of unauthorized data usage or disclosure. Within this framework, data holders are discouraged from using these protective measures to restrict users from effectively sharing data with third parties. In cases where data is acquired deceitfully by a recipient and used for unauthorized purposes, the Regulation insists on the immediate destruction of such data and a cessation of its distribution. This rule safeguards the integrity of the data-sharing environment and enforces adherence to legal norms within the digital sphere.

Chapter IV - Unfair terms related to data access and use between enterprises Another safeguard to the integrity of the data market is put in place by Art. 13 by defining what terms can be considered unfair related to data access and use between enterprises. Contractual terms that have as object or effect to (a) limite or exclude the liability of the party imposing the term for intentional acts or gross negligence; (b) exclude remedies in case of non-performance or breach of obligations (c) give the imposing party the exclusive right of judging whether the data supplied are in conformity with the contract or to interpret any term of the

[28] Art. 5(8)

[29] https://www.europarl.europa.eu/RegData/commissions/itre/inag/2023/07-14/ITRE$_A$G(2023)751822$_E$N.pdf

contract[30] (d) inappropriately limit the remedies (e) allow the imposing party to access and use data in a manner that violates significantly the legitimate interest of the other contracting party (f) prevent a party from having a copy of the data, or a disproportionate limit on the use of such data (g) prevent the termination of a contract on an unreasonably short notice[31] In case of a contract containing such unfair terms, it is to be consider invalid with regards to those part, while the rest remains legally binding.

Chapter VI - Switching between data processing services Providers of data processing services shall remove any obstacles which prevents the customers from (a) terminating the contractual agreement (b) concluding new agreements with a different provider (c) porting data to another provider (d) maintaining functional equivalence of the service

Furthermore, this Regulation sets minimum standards for contractual terms about switching providers mandating the contractual feasibility of switching, while providing assistance during the process and ensuring continuity of service, exhaustive specification of all data and applications exportable and a minimum period for data retrieval of 30 days. For what concerns switching charges, the general guideline mirrors that of the charges on data availability: they 'shall not exceed the costs incurred by the provider of data processing services that are directly linked to the switching process concerned'.[32]

One major aspect of this Act revolves around interoperability. While if a provider is offering Infrastructure-as-a-Service (IaaS), i.e. its service revolves around landing computing power, networks, servers, etc., it has to ensure the customers enjoys a switching process without excessive hurdles and functional equivalence at the new provider, if the provider falls under a different category of data processing service, it shall • make open interfaces available to the public for free, and • ensure its services are compatible with existing open interoperability standards. If no open interoperability specifications or European standards are available, the provider shall facilitate the export of all data and metadata in a machine readable format.

These rules follow the spirit of Commission's vision of data sharing market that is free to develop and at the same time safe for the consumers and for the smaller players. It encourages a healthy competitive environment where there is minimal friction in

[30]Details on dispute settlements are found in Art. 10 which specifies that those bodies shall be impartial, independent, possess the necessary expertise, and able to swiftly solve cases. That article also provides for Member States to establish one if there is no body that meets the described criteria.

[31]On the Provisional Agreement, there is also a new addition that considers unfair term the ability to 'substantially alter the price stipulated in the contract or any other substantive conditions'.https://www.europarl.europa.eu/RegData/commissions/itre/inag/2023/07-14/ITRE$_A$G(2023)751822$_E$N.pdf

[32]Art. 25

navigating data services. The regulation promotes openness, interoperability, and customer freedom, encouraging a healthy competitive environment where data can be shared easily and services can be switched with minimal friction.[33]

Chapter VIII - Interoperability. Despite being almost left out of this regulation, data spaces are reserved a spot in the interoperability section with the idea of developing the legislative framework with other Regulations or Directives.[34]. Data spaces are required to clearly explain some information, like data structures, formats, and related vocabularies, alongside inherent classification schemes and taxonomie, as well as licensing and data collection methodology, in order to facilitate a comprehensive understanding for the data recipient, enabling them to optimally find, access and utilize the data. Furthermore, there is an explicit reference to application programming interfaces (APIs) that serve as a mean to interact with the data, which should 'enable automatic access and transmission of data between parties'.[35] The provision then shifts to operational interoperability of smart contracts highlighting the necessity to adopt services that incorporate this instruments seamlessly, promoting smooth transaction and agreements.

Attention is dedicate also to this topic of smart contracts that would have to be robust, provide the capability of safe termination and interruption, data archiving and continuity, and most importantly access control. This last characteristic is crucial because it helps solve the issues of trust and authentication in data sharing, and it's application to IoT devices has been widely studied.[36]

Conclusions. Ultimately, the Data Act appears to be somewhat premature and insufficient, given the extensive amount of forthcoming legislation that remains to be formulated and ratified through both implementing and delegated acts, as well as through the standard legislative process.[37]. Notably, the idea of data spaces seems to be in nascent stages, lacking substantial development. The regulatory frameworks are being established even before the inception of a unified European Data Space. However, the crux of the issue transcends lies in the apparent disparity between the efforts and resources channelled into devising regulations, as opposed to those invested in nurturing and expanding the subject of matter itself.

[33]Art. 26

[34]At the moment of writing, there are a couple of Regulations being discussed on a Common Data Space for Health

[35]Art. 28

[36]See Yuanyu Zhang et al. (2018). "Smart contract-based access control for the internet of things". In: *IEEE Internet of Things Journal* 6.2, pp. 1594–1605 for a model of multiple data access smart contracts managing IoT devices.

[37]As an example, Art. 28(5), Art. 29(5), Art. 30(6) on the adoption of standards in the case data spaces requirements, data processing services and smart contracts

4.3.2 Portability

Data portability is one of the recurring themes in EU legislation on data, the GDPR already provides for a 'right to data portability' with which data subjects can obtain their data from a vendor and transfer them to a different provider. The DA aim to expand and detail this right, since while the GDPR allows for direct data transfer between controllers when technically possible, the GDPR doesn't require system compatibility. In fact, it states 'where technically feasible', the data subject shall have the right to have the personal data transmitted directly from one controller to another, and Rec. 68 of the GDPR specifies that it does not impose an "obligation for the controllers to adopt or maintain processing systems which are technically compatible". The technical requirements for implementing data portability might be very high since, ideally and de facto, a technical measure should be established that facilitates data transfers, which seems to be at odds with the provision of Recital 68. The right to data portability as far as it concerns the reception of a user's personal data (not the direct data transfer to another platform) applies to all automated processing systems independent of whether transfer of data is already technically possible.[38] Still, data controllers would have to implement processes for handling and documenting user's requests. In order to enable data portability, platform operators will need to use 'structured, commonly used and machine-readable' data formats and templates. In practice, the receiving service should be able to process the data extracted from another platform in an efficient manner. This might force some platform providers to change their design to a certain standard, which still needs to be established.[39]

However, it has to be kept in mind that data portability is a subset, and not equivalent to interoperability. Interoperability between online social networks for example would enable users to connect with each other irrespective of their social network affiliation. Facebook users would then be able to directly post a message on someone's Google+ page. With data portability, by contrast, Facebook users could take their profile and message history to Google+ and open a new Google+ account based on this information.[40] In other words, they would not have to start from scratch when changing or additionally using platforms. The right to data portability would, to give another example apart from social networks, also imply that users

[38]Inge Graef, Jeroen Verschakelen, and Peggy Valcke (2013). "Putting the right to data portability into a competition law perspective". In: *Law: The Journal of the Higher School of Economics, Annual Review*, pp. 53–63.

[39]Christopher S Yoo (2011). "When antitrust met Facebook". In: *Geo. Mason L. Rev.* 19, p. 1147.

[40]Inge Graef (2015). "Mandating portability and interoperability in online social networks: Regulatory and competition law issues in the European Union". In: *Telecommunications Policy* 39.6, pp. 502–514.

could move their information uploaded into a cloud storage service directly to a competing cloud storage service, which is one of the main targets envisioned in the DA.

When it comes to legal aspects, data portability is likely to raise issues like privacy and data security. If data is portable, a single identity fraud can turn into a long-lasting breach of personal data, since a hacker can easily port his false identity to many other platforms. This is way platform providers will likely have to expand their investment in data security measures. Another legal issue raised by data portability is the fact that multiple individuals might claim control over certain information. Making this information portable might infringe property rights. The right to data portability is also difficult to apply in cases where multiple data subjects are involved who disagree on the data transfer. For example, multiple people might appear in a photograph, allowing one user to transfer a second user's information may violate the privacy rights of second user. What is more, people can easily evade privacy restrictions placed by the initial platform by porting the data over to another platform not subject to these restrictions. These legal uncertainties will once again impose challenges to platform providers.[41]

Switching costs are fueled by direct network effects and coordination costs. Without data portability, contacts cannot be transferred to another platform and information that has once been shared, i.e. data that the user has directly or indirectly 'invested', such as messages, photos, reputation and search histories, remain with the original platform. The user is therefore more likely to stay with the platform that he/she initially provided his data to, although rival platforms might otherwise be more attractive to him/her. This might harm competition since potential competitors might not have an incentive to innovate and offer better services, knowing that users will nevertheless remain with the incumbent platform.[42]

If an user comes from system A which has a certain format and structure, system B, the new service, has to adapt itself and it can cause issues to the subject that now has no indication and possibility to view the data the way he/she was used to. If data portability is not guaranteed, platform A can potentially preclude platform B from entering the market or from gaining a higher market share. This is possible since users can only switch between the platforms at high costs if they cannot take their data with them. They might not even switch to platform B although it might be significantly more attractive to them because of prohibitive switching costs – they are locked-in. The lock-in caused by this behavior from tech companies leveraging the network effect, is evident in everyday life and severely hinders competition. Lock-in

[41] Barbara Engels (2016). "Data portability among online platforms". In: *Internet Policy Review* 5.2.

[42] Ibid.

and network effect work together to create the perfect storm where once a piece of technology, a service, is adopted by a large group, the cost of switching becomes troublesome enough that both the users and the market freeze.[43]

If that happens, only an extreme event can move the situation from the impasse, either a sudden change in users' sentiment toward the service, a disruptive legislative intervention or, more rarely, a disruptive competitor providing a different system.

4.4 An opinion from IDSA and SalusCoop

The International Data Spaces Association (IDSA), represented by a team of experts with different professional backgrounds, has provided me with a nuanced understanding of the data sharing market's current state. Söntje Hilberg, head of the legal department, focuses on the path followed at the European level. Despite significant advancements, harmonizing new rules with member states' laws presents a formidable challenge, as well as dealing with corporate fatigue from overwhelming regulatory compliances.

Trust surfaces as a central theme in IDSA's discourse, perceived as a critical connector in the data sharing ecosystem. The experts argue for a synergistic approach, where transparency, adherence to common rules, and open-source commitments cultivate trust. However, the pathway to materializing trust necessitates more pragmatic standards and guidelines, fostering a milieu where trust is not only nurtured but also operable.

The European Data Innovation Board emerges as a potent entity in this context, offering promising avenues for fostering trust and collaboration. The board is anticipated to spearhead efforts in defining practical standards, thereby paving the way for a data sharing environment characterized by reliability and mutual benefit.

Situated in Catalunya, SalusCoop Cooperative epitomizes a citizen-centric approach to data management. As a non-profit entity, it empowers citizens to have a decisive role in managing and sharing their personal data, embodying a cooperative spirit that harmonizes individual autonomy with communal welfare.

Trust holds paramount significance in the cooperative's operational philosophy. Through a robust framework featuring methodologies like blockchain and zero-knowledge, SalusCoop Cooperative guarantees data anonymity and integrity, weaving trust intricately into its operational fabric. This initiative demonstrates how fostering trust transcends theoretical discourse, translating into tangible, solution-oriented practices.

[43] Joseph Farrell and Paul Klemperer (2007). "Chapter 31 Coordination and Lock-In: Competition with Switching Costs and Network Effects". In: ed. by M. Armstrong and R. Porter. Vol. 3. Handbook of Industrial Organization. Elsevier, pp. 1967–2072.

The convergence of insights from IDSA and the Salus Cooperative paints a hopeful picture of the future of personal data management. It echoes a symbiotic relationship, where the strengths of international data spaces and data cooperatives coalesce to forge a resilient, trust-centric data governance ecosystem.

To harness the full potential of this synergy, amplifying the visibility of data cooperatives and data spaces becomes indispensable. Comprehensive public awareness campaigns and educational initiatives could illuminate the multifaceted benefits these cooperatives offer, emphasizing their role as catalysts for fostering a user-centric, collaborative data governance model.

The commonality of intent between data cooperatives and data spaces foretells a promising narrative for data management, both personal and industrial. This synergy hints at the possibility of a change in the data environment where collaboration, trust, and mutual benefit are at the forefront. As the European landscape evolves, fostering education, understanding, and trust in the sector becomes imperative, casting data cooperatives and data spaces as the torchbearers of this transformative voyage.

An insight that could be gained from these two conversation is that there is more work to be done in terms of adapting the public sector to standards that are already implemented by private companies, that there has to be an alignment of interests of the Member States and the EU because collaboration will build the foundation for more awareness, and consequently a wider adoption of models like data cooperatives and data spaces. And cooperation is needed also to pool funds to destinate towards projects like data cooperatives, which are non-profit and help competitiveness in the market trying to disrupt the model of gatekeepers hoarding and centralizing enormous amounts of data. If the public is not aware of the possibilities that already exist in the market and funds are not provided, all the work done with the various pieces of legislation will be severely hindered.

Conclusions

5.1 Discussion

The DGA represents a part of the regulatory framework overseeing a subset of data intermediaries, who find themself under the scrutiny. One significant concern is the high degree of legal uncertainty is brings because of the vagueness in the definition of the scope of the regulation, testified by the fact that different bodies are entitled to intervene in the matter: competent authorities in Member States, the Commission and the EDIB. This uncertainty targets not only established businesses, but also potentially dissuades the entrance of new players in the market, thus creating the paradox of an act made to foster development, yet with the tangible potential to hindering it.[1] This sentiment is further fueled by the perceived shortcomings in the act's ability to effectively address some data sharing issues, such as data quality and provenance, enforcement of purpose limitation of data use, and the lack of standardized protocols. Furthermore, some of these issues are not directly tackled, like data quality and provenance, which are left to the operators to deal with, the others are inconclusively referred to, waiting for bodies or courts to be decided. The uncertainty brought by this lack of precision in definition adds unpredictability in a market that would need a clear and precise set of rules, while also avoiding being to interventionist. Striking a balance between these two needs is vital for regulation in any sector, and even more in one that is in its developing stages. Another friction point is the need for structural separation of services: while it is necessary in some situation concerning competition law, as in markets where the threat of cross-data usage is highest.[2] This requirement also clashes against the possible disadvantages of data intermediation services against unregulated or less regulated ones[3].

On this topic Baloup et al. (June 2021) argue that data neutrality is an excessive measure considering the market situation of data intermediation services is not comparable to the one of gatekeepers as described in the DMA, thus may causing

[1] Heiko Richter (2023). "Looking at the Data Governance Act and Beyond: How to Better Integrate Data Intermediaries in the Market Order for Data Sharing". In: *GRUR International*, ikad014.

[2] Julie Baloup et al. (June 2021). *White Paper on the Data Governance Act.*

[3] Inge Graef and Raphael Gellert (2021). In.

less innovation. Since this Regulation is a novelty in terms of theme, as it approaches a topic that has no precedents in EU legislation, the true outcome is yet to be seen, and for now, there can only be speculation on its real effectiveness. It is significant that Art. 35 includes the requirements of a ex-post evaluation of this legislation that 'shall be accompanied, where necessary, by legislative proposals' to be carried out by 24 September 2025, even though it refers only to the compliance aspects, and not to the effectiveness the Act. It explicitly mentions the 'application and functioning of the rules on penalties', the level of enforcability of such penalties and the level of compliance by actors not established in the Union to this Regulation.[4]

While all this issues could be classified as speculation, the handling of personal data by data intermediaries clashes with the obligations of the GDPR, since there is no exception for specific types of intermediaries such as data cooperatives, or data trustee which often handle this kind of data. The latter in particular play a significant role in helping data subjects assert their rights, whether for pseudonymisation processes or as agents mandating data protection preferences. This is true specifically for Personal Information Management Systems (PIMS), in some ways similar to PDS, which are used not only manage consent but can extend their services to enforce data subjects' rights and eventually claim damages in cases of violations of data protection norms. More worrying is the fact that the DGA is more stringent than the GDPR in the context of data processing for a different purpose than initially outlined, provided the data subject gives their consent: the GDPR allows for it, the DGA, Art. 12(a), explicitly forbids such actions.

This legislative landscape has spurred a discussions on the feasibility and 'friendliness' of the existing data protection law towards intermediaries. Several critical issues are central to this conversation: firstly, the extent to which data protection laws empower intermediaries to manage data subjects' consent adequately, considering the strong rights vested in data subjects by the GDPR. Secondly, the capability of data trustees to exercise data subjects' rights to rectification and erasure as outlined in the GDPR. Thirdly, the broader legal framework concerning the liability of data trustees, especially in cases of breach of data protection law. Lastly, the ongoing debate focuses on the potential contractual limitations that data holders might impose on mandating data trustees, and the repercussions this could have on competition and market entry for new trustees.[5]

In response to these complex issues, suggestions for reform have been prolific. While some argue for clear legislative action to delineate trustee obligations, quality

[4]The only topic not related to compliance refers to type of data altruism organization and which kind of data are shared through them.

[5]Heiko Richter (2023). "Looking at the Data Governance Act and Beyond: How to Better Integrate Data Intermediaries in the Market Order for Data Sharing". In: *GRUR International*, ikad014.

standards, and liability questions, others propose relying on further guidance from supervisory institutions, expressing skepticism towards additional sector-specific regulations parallel to the GDPR.[6] Regardless, the urgency for more integrated policy interventions is necessary to provide a fruitful environment for data intermediaries within the market structure for data sharing, without undermining the existing data protection statutes.

Data spaces, despite being reserved a place in both regulation, are not their main focus, but will still be impacted for what concerns guidelines to follow. At the moment, the only ongoing procedure is at its initial stages when the Commission has received all the feedback and the is no written proposal.[7]

In examining the Data Act, it's noteworthy that the proposal doesn't reference data intermediation services, especially given that the DGA is viewed as the basis for for the Data Act.[8] Data intermediaries have the potential to operationalize the data access rights outlined in Chapter III of the Regulation. They could substantially decrease transactional expenses, thereby facilitating widespread data-driven innovation. This is because they might possess a deeper comprehension of and capability to amalgamate third-party data demands for innovation, as well as efficiently consolidate data from diverse sources. Additionally, they can further refine user data to cater to distinct data recipients' needs and oversee data transfers from both technical and legal standpoints. Yet, the Data Act overlooks such potentials.[9]

A critical discussion revolves around the extent to which the Data Act permits users to monetize the data they obtain under Art. 4 or distribute to third parties via Art. 5. Essentially, this delves into whether a user could solely profit from data without direct service benefits. In such scenarios, data intermediaries are uniquely positioned to provide substantial remuneration and then redistribute the acquired data. However, Art. 6(2)(c) seems to oppose such commercialization. The article suggests that data intermediaries can't merely offer payments to users for data, which they can then repurpose and offer to third parties for innovation.[10]

This thesis had the objective to give an overview on these two new piece of legislation and try to check whether the initial proposition of the Commission of boosting the data sharing sector has been followed. It is clear that the addressing

[6]Heiko Richter (2023). "Looking at the Data Governance Act and Beyond: How to Better Integrate Data Intermediaries in the Market Order for Data Sharing". In: *GRUR International*, ikad014.

[7]https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12491-Data-sharing-in-the-EU-common-European-data-spaces-new-rules-$_en$

[8]European Commission (2020). *COM(2020) 66 final, A European strategy for data.*

[9]Josef Drexl et al. (2022). "Position Statement of the Max Planck Institute for Innovation and Competition of 25 May 2022 on the Commission's Proposal of 23 February 2022 for a Regulation on Harmonised Rules on Fair Access to and Use of Data (Data Act)". In: *Max Planck Institute for Innovation & Competition Research Paper* 22-12.

[10]Ibid.

ambiguity should be the priority when implementing these regulations and drawing up guidelines. Other stimula to research on interoperability will have to wait until the provisions of these regulations have their effect because there is already a substantive corpus of literature on the subject. Further research on the topic should focus on data altruism organization, left aside here, and dwell more into data cooperatives because they represent an opportunity to see a different model of data exchange which is primarely in the hands of the data subject.

5.2 Conclusions

In the end, the EU is demonstrating once again to be at the forefront of the battle to guarantee its citizens adequate safeguards to the use of their data. It also proves that the EU is conscious of the issues in the European data economy and is trying to balance the interests of economic development and citizens safety. But despite its best intentions, the final product, if it can be call that, seen the further interventions needed, is problematic in many ways. The data sharing economy is still not developed, its causes are known and tacking these problems has been one of the objective of EU action in the last decade, and even more with Von Der Leyen's Commission. From its work, the Digital Market Act, Digital Service Act and Data Governance Act have seen the light of day, and the Data Act will soon. Whether the last two Regulation will provide the boost needed in the sharing sector to thrive is still to be seen, but with this preliminary work, I think there is a fair possibility that the desired objective will be achieved and the corrections that will arrive from the inputs of the EDIB and from the practices of competent bodies will determine the final judgement on these Regulations.

Bibliography

Alaimo, Cristina and Jannis Kallinikos (2022). "Organizations decentered: Data objects, technology and knowledge". In: *Organization Science* 33.1, pp. 19–37.

Auer, Dirk and Nicolas Petit (2015). "Two-sided markets and the challenge of turning economic theory into antitrust policy". In: *The Antitrust Bulletin* 60.4, pp. 426–461.

Baloup, Julie et al. (June 2021). *White Paper on the Data Governance Act.*

Basin, David, Søren Debois, and Thomas Hildebrandt (2018). "On purpose and by necessity: compliance under the GDPR". In: *Financial Cryptography and Data Security: 22nd International Conference, FC 2018, Nieuwpoort, Curaçao, February 26–March 2, 2018, Revised Selected Papers 22*. Springer, pp. 20–37.

Commission, European (2020). *COM(2020) 66 final, A European strategy for data.*

— (2022). *Commission Staff Working Document. Impact assessment report accompanying the document Proposal for a Regulation of the European Parliament and of the Council on harmonized rules on fair access to and use of data (Data Act).*

Commission, European and Everis (2018). *Study on data sharing between companies in Europe.*

Curry, Edward, Simon Scerri, and Tuomo Tuikka (2022). *Data Spaces: Design, Deployment and Future Directions.*

Davenport, Thomas H, Laurence Prusak, et al. (1998). *Working knowledge: How organizations manage what they know*. Harvard Business Press.

De Reuver, Mark, Carsten Sørensen, and Rahul C Basole (2018). "The digital platform: a research agenda". In: *Journal of information technology* 33.2, pp. 124–135.

Delors, Jaques (1985). *The Dignity of Work and the Reasons for Peace.*

Drexl, Josef et al. (2022). "Position Statement of the Max Planck Institute for Innovation and Competition of 25 May 2022 on the Commission's Proposal of 23 February 2022 for a Regulation on Harmonised Rules on Fair Access to and Use of Data (Data Act)". In: *Max Planck Institute for Innovation & Competition Research Paper* 22-12.

Elmasri, Ramez, Sham Navathe, et al. (2014). *Fundamentals of database systems*. Vol. 7. Pearson.

Engels, Barbara (2016). "Data portability among online platforms". In: *Internet Policy Review* 5.2.

Evans, David S and Richard Schmalensee (2013). *The antitrust analysis of multi-sided platform businesses*. Tech. rep. National Bureau of Economic Research.

Farrell, Joseph and Paul Klemperer (2007). "Chapter 31 Coordination and Lock-In: Competition with Switching Costs and Network Effects". In: ed. by M. Armstrong and R. Porter. Vol. 3. Handbook of Industrial Organization. Elsevier, pp. 1967–2072.

Franklin, Michael, Alon Halevy, and David Maier (Dec. 2005). "From Databases to Dataspaces: A New Abstraction for Information Management". In: *SIGMOD Rec.* 34.4, pp. 27–33.

Ghazawneh, Ahmad and Ola Henfridsson (2015). "A paradigmatic analysis of digital application marketplaces". In: *Journal of Information Technology* 30.3, pp. 198–208.

Gobbato, Sara (2020). "Open Science and the reuse of publicly funded research data in the new Directive (EU) 2019/1024". In: *Journal of Ethics and Legal Technologies* 2.2.

Graef, Inge (2015). "Mandating portability and interoperability in online social networks: Regulatory and competition law issues in the European Union". In: *Telecommunications Policy* 39.6, pp. 502–514.

Graef, Inge and Raphael Gellert (2021). In.

Graef, Inge, Jeroen Verschakelen, and Peggy Valcke (2013). "Putting the right to data portability into a competition law perspective". In: *Law: The Journal of the Higher School of Economics, Annual Review*, pp. 53–63.

Halevy, Alon, Michael Franklin, and David Maier (2006). "Principles of Dataspace Systems". In: *Proceedings of the Twenty-Fifth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*. PODS '06. Chicago, IL, USA: Association for Computing Machinery, pp. 1–9.

Hardjono, Thomas and Alex Pentland (2019). *Data Cooperatives: Towards a Foundation for Decentralized Personal Data Management*.

Hildebrandt, Mireille (2020). *Law for computer scientists and other folk*. Oxford university press.

Ichihashi, Shota (2021). "Competing data intermediaries". In: *The RAND Journal of Economics* 52.3, pp. 515–537.

Jacobsen, Annika et al. (2020). *FAIR principles: interpretations and implementation considerations*.

Lehtiniemi, Tuukka (2017). "Personal data spaces: An intervention in surveillance capitalism?" In: *Surveillance & Society* 15.5, pp. 626–639.

Lenzerini, Maurizio (2002). "Data integration: A theoretical perspective". In: *Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pp. 233–246.

Mărcuț, Mirela (2017). *Crystalizing the EU digital policy: an exploration into the digital single market*. Springer.

Pizzetti, Franco (2016). *Privacy e il diritto europeo alla protezione dei dati personali: Dalla Direttiva 95/46 al nuovo Regolamente europeo*. G. Giappichelli Editore.

Richter, Heiko (2023). "Looking at the Data Governance Act and Beyond: How to Better Integrate Data Intermediaries in the Market Order for Data Sharing". In: *GRUR International*, ikad014.

Richter, Heiko and Peter R Slowinski (2019). "The data sharing economy: on the emergence of new intermediaries". In: *IIC-International Review of Intellectual Property and Competition Law* 50, pp. 4–29.

Rochet, Jean-Charles and Jean Tirole (2003). "Platform competition in two-sided markets". In: *Journal of the european economic association* 1.4, pp. 990–1029.

— (2006). "Two-sided markets: a progress report". In: *The RAND journal of economics* 37.3, pp. 645–667.

Rysman, Marc (2009). "The economics of two-sided markets". In: *Journal of economic perspectives* 23.3, pp. 125–43.

Scerri, Simon et al. (2022). "Common European Data Spaces: Challenges and Opportunities". In: *Data Spaces: Design, Deployment and Future Directions*, pp. 337–357.

Somaini, Laura (2020). "Regulating the dynamic concept of non-personal data in the EU: from ownership to portability". In: *Eur. Data Prot. L. Rev.* 6, p. 84.

Tiwana, Amrit, Benn Konsynski, and Ashley A. Bush (2010). "Research Commentary—Platform Evolution: Coevolution of Platform Architecture, Governance, and Environmental Dynamics". In: *Information Systems Research* 21.4, pp. 675–687.

Tuomi, Ilkka (1999). "Data is more than knowledge: Implications of the reversed knowledge hierarchy for knowledge management and organizational memory". In: *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. 1999. HICSS-32. Abstracts and CD-ROM of Full Papers*. IEEE, 12–pp.

Wang, Yihan, Shaoxu Song, and Lei Chen (Sept. 2016). "A Survey on Accessing Dataspaces". In: *SIGMOD Rec.* 45.2, pp. 33–44.

Wernick, Alina (2020). "Defining Data Intermediaries: A Clearer View Through the Lens of Intellectual Property Governance". In: *Technology and Regulation* 2020, pp. 65–77.

Wilkinson, Mark D et al. (2016). "The FAIR Guiding Principles for scientific data management and stewardship". In: *Scientific data* 3.1, pp. 1–9.

Yoo, Christopher S (2011). "When antitrust met Facebook". In: *Geo. Mason L. Rev.* 19, p. 1147.

Zhang, Yuanyu et al. (2018). "Smart contract-based access control for the internet of things". In: *IEEE Internet of Things Journal* 6.2, pp. 1594–1605.