



European
Commission

JRC SCIENCE FOR POLICY REPORT

Mapping the landscape of data intermediaries

*Emerging models for
more inclusive data governance*

Micheli, M.
Farrell, E.
Carballa-Smichowski, B.
Posada-Sánchez, M.
Signorelli, S.
Vespe, M.

2023

This publication is a Science for Policy report by the Joint Research Centre (JRC), the European Commission's science and knowledge service. It aims to provide evidence-based scientific support to the European policymaking process. The contents of this publication do not necessarily reflect the position or opinion of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication. For information on the methodology and quality underlying the data used in this publication for which the source is neither Eurostat nor other Commission services, users should contact the referenced source. The designations employed and the presentation of material on the maps do not imply the expression of any opinion whatsoever on the part of the European Union concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

EU Science Hub

<https://joint-research-centre.ec.europa.eu>

JRC133988
EUR 31576 EN

Print ISBN 978-92-68-05063-7 ISSN 1018-5593 doi:10.2760/8943 KJ-NA-31-576-EN-C
PDF ISBN 978-92-68-05064-4 ISSN 1831-9424 doi:10.2760/261724 KJ-NA-31-576-EN-N

Luxembourg: Publications Office of the European Union, 2023

© European Union, 2023



The reuse policy of the European Commission documents is implemented by the Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39). Unless otherwise noted, the reuse of this document is authorised under the Creative Commons Attribution 4.0 International (CC BY 4.0) licence (<https://creativecommons.org/licenses/by/4.0/>). This means that reuse is allowed provided appropriate credit is given and any changes are indicated.

For any use or reproduction of photos or other material that is not owned by the European Union permission must be sought directly from the copyright holders.

How to cite this report: Micheli, M., Farrell, E., Carballa-Smichowski, B., Posada-Sánchez, M., Signorelli, S., Vespe, M., *Mapping the landscape of data intermediaries — Emerging models for more inclusive data governance*, Publications Office of the European Union, Luxembourg, 2023, doi:10.2760/261724, JRC133988.

Layout: Carmen Capote de la Calle

SCIENCE FOR POLICY REPORT

Mapping the landscape of data intermediaries

*Emerging models for
more inclusive data governance*

Micheli, M.
Farrell, E.
Carballa-Smichowski, B.
Posada-Sánchez, M.
Signorelli, S.
Vespe, M.

2023

Table of contents

- 05 Abstract
- 05 Acknowledgements
- 07 Executive summary
- 10 **1. Introduction**
 - 1.1 EU policy context
 - 1.2 What you will find in this report
- 17 **2. Towards more inclusive data governance**
 - 2.1 Data governance
 - 2.2 Inclusive data governance
 - 2.3 Data intermediaries
- 32 **3. The economic sustainability of data intermediaries**
 - 3.1 Defining a business model
 - 3.2 Conditions of sustainability of data intermediaries
- 40 **4. The landscape of data intermediaries**
 - 4.1 Personal Information Management Systems (PIMS)
 - 4.2 Data cooperatives
 - 4.3 Data trusts
 - 4.4 Data unions
 - 4.5 Data marketplaces
 - 4.6 Data sharing pools
- 62 **5. Concluding remarks**
- 66 References
- 72 List of abbreviations and definitions
- 73 List of boxes
- 73 List of figures
- 73 List of tables

Abstract

This report provides a landscape analysis of key emerging types of data intermediaries. It reviews and synthesises current literature in this area, with the goal of identifying shared elements and definitions. Its main objective is to contribute to the establishment of a common vocabulary concerning data intermediaries among EU policy makers, experts, and practitioners. Six types of data intermediaries are presented in detail: Personal Information Management Systems (PIMS), data cooperatives, data trusts, data unions, data marketplaces, and data sharing pools. For each type, the report provides information about how the intermediary works, its main features, selected examples, and business model considerations. The report is grounded in multiple perspectives spanning sociological, legal, and economic disciplines. In particular, the analysis is informed by the concept of inclu-

sive data governance, situated in the context of the recent Data Governance Act, and problematised according to the economic and management literature.

The report findings highlight the fragmentation and heterogeneity of the field. Data intermediaries can range from individualistic and business-oriented types to more collective models that support greater public engagement in data governance, including by underprivileged actors; while certain types of data intermediaries aim at facilitating economic transactions between data holders and users, others mainly seek to produce collective benefits or public value. In the conclusions, the report presents a series of take-aways regarding the main obstacles that may be faced by data intermediaries and suggests possible future research directions for empirical work in this field.



Acknowledgements

The authors would like to express their gratitude to all the experts that participated in the online workshop ‘Understanding data intermediaries for more inclusive data governance’ (held on March 29th 2023), for sharing their knowledge and insights about the topics addressed in the report. The authors are particularly grateful to the following workshop participants who also provided written feedback on an early draft of the report and significantly helped to improve and consolidate its content: Sylvie Delacroix, Paolo Gerli, Inge Graef, Heleen Jansen, Viivi Lähteenoja, Manuel Portela Char-

nejovsky, Barbara Prainsack, Maria Savona, and Lukas von Ditfurth.

The authors would like to express their sincere gratitude to colleagues from the Data Policy and Innovation Unit of the European Commission Directorate General for Communications Networks, Content and Technology (DG-CNECT) for their valuable feedback, continuous support and insights in relation to the legal texts underpinning this report: Katja Vaahtera, Malte Beyer-Katzenberger, Federico Milani, Anna Ludin, Maria Rosaria Coduti, Justyna Palczynska and Narcisa Elena Toderas-Alexa.

A special thanks also goes to Carlos Torrecilla-Salinas for supporting this project while acting as Head of Unit at the JRC Digital Economy Unit, as well as to colleagues Brooke Tapsall, Ana Fernández Cruzado, Alexander Kotsev and Micheal Lutz (while deputy Head of Unit) who provided invaluable help along the road that led to this publication. In addition, the Unit's Communication team also provided significant support, in particular Carmen Capote de la Calle, for the design of the report, and Miriam Giubilei and Francesca Siciliano, for coordinating the dissemination activities.

The authors would like to thank Koen Jonkers (Editor in Chief of the JRC Editorial Review Board), Lucia Vesnić-Alujević and Charles Heard, whose reviews of the report helped the authors improve its content and format.

Finally, an important recognition goes to Jaime Bernal Cereceda and Alessio Tartaro, who joined the team as trainees during the first phases of the project, and significantly contributed to shaping the initial direction of the report.

Authors

- **Marina Micheli**
European Commission, Joint Research Centre, Ispra, Italy
- **Eimear Farrell**
European Commission, Joint Research Centre, Ispra, Italy
- **Bruno Carballa-Smichowski**
European Commission, Joint Research Centre, Seville, Spain
- **Monica Posada-Sánchez**
European Commission, Joint Research Centre, Ispra, Italy
- **Serena Signorelli**
European Commission, Joint Research Centre, Ispra, Italy
- **Michele Vespe**
European Commission, Joint Research Centre, Ispra, Italy

Contributors

- **Sylvie Delacroix**
University of Birmingham
- **Paolo Gerli**
Edinburgh Napier University
- **Inge Graef**
Tilburg University
- **Heleen Janssen**
University of Cambridge
- **Viivi Lähteenoja**
MyData Global
- **Manuel Portela Charnejovsky**
Universitat Pompeu Fabra
- **Barbara Prainsack**
University of Vienna
- **Maria Savona**
University of Sussex and Luiss
- **Lukas von Ditfurth**
University of Passau

The views expressed are purely those of the authors and may not in any circumstances be regarded as stating an official position of the European Commission. All errors remain the sole responsibility of the authors.

Executive summary

Data intermediaries are nascent organisations and knowledge about them is still developing. This report aims at increasing understanding about data intermediaries by providing a landscape analysis of key emerging types. It reviews and synthesises current academic and policy literature on this topic with the goal of identifying common elements and definitions for data intermediaries. One of its main objectives is to contribute to the establishment of a common vocabulary in this area among EU policy makers, experts, and practitioners.

The report is the first outcome from a series of science for policy activities that the JRC is conducting to support the implementation of the Data Governance Act. It is the result of a desk review of the main scholarly and policy literature, and it is enriched by feedback from the European Commission officials working on data policy, as well as from academics and experts external to the Commission, who were involved through a dedicated online workshop.

Perspectives

The review of data intermediaries featured in this report is grounded in multiple perspectives spanning sociological, legal, and economic disciplines. More precisely, the analysis is: 1) informed by the concept of inclusive data governance; 2) contextualized in the recent Data Governance Act; and 3) problematised according to the economic and management literature on business models. We briefly introduce the perspectives here:

- 1. The notion of inclusive data governance** constitutes the conceptual underpinning of the report and is

adopted to articulate a widespread concern around the unbalanced data practices that characterise the current digital ecosystem, as well as to emphasise the urgency of exploring alternative data practices to overcome these. [Section 2](#) defines the concept and explains how data intermediaries might contribute, at least to some degree, to address the power asymmetries of the current data society and economy. Data intermediaries in fact, might allow broader stakeholder participation in decision-making concerning data access, control, sharing and use, allowing both economic entities and individual data subjects to have greater agency. Furthermore, as enablers of data sharing, data intermediaries might also foster the production of greater economic and social value from data. Both aspects — stakeholder participation and value production and distribution — are important steps for building a fairer data ecosystem.

- 2. The Data Governance Act (DGA)** is the first major legislative initiative implementing the European Strategy for Data. It entered into force in June 2022 and will be applicable from September 2023. The DGA is a cross-sectoral instrument that aims at increasing trust in voluntary data sharing for purposes of both economic growth and public interest, by creating an enabling framework for different data use and reuse situations. Notably, it aims to boost trustworthy data sharing by regulating data intermediation services providers (DISPs),



a specific type of data intermediary. This report aims to clarify which kind of data intermediaries are covered by the DGA ([Section 2.3](#)) and elucidates how such services relate to the broader landscape of data intermediaries for inclusive data governance presented in [Section 4](#).

3. The report examines the concept of **business models in relation to data intermediaries**, in order to analyse the strategies and conditions that could make such entities economically sustainable. Reaching economic sustainability is one of the main challenges currently faced by data intermediaries to become relevant actors in the data economy and to deliver the expected positive impacts anticipated both by the academic literature and by policy makers. Therefore, [Section 3](#) of the report introduces key issues at stake for economic sustainability, while business model considerations for each type of data intermediary are included in [Section 4](#).

Types of data intermediaries

The report describes **six types of data intermediaries**: Personal Information Management Systems (PIMS), data cooperatives, data trusts, data unions, data marketplaces and data sharing pools ([Section 4](#)). These differ according to various parameters, yet they are also not completely distinct from one another and might overlap in certain aspects. For each type of data intermediary that is presented, the report provides information about how it works, its main features, selected examples, and business model considerations. Finally, it summarises the main characteristics in a comprehensive table ([Table 2](#)). The table

provides a concise summary of the key features of these six types of data intermediaries what are commonly mentioned in this field, and which can potentially foster more inclusive data governance in Europe and beyond.

This analysis presented in this report draws connections between the wider landscape of data intermediaries for inclusive data governance and the data intermediation services providers (DISPs) covered by the current EU regulatory framework. Besides describing six popular types of data intermediaries, it elucidates what DISPs are (i.e., the specific types of data intermediaries that will have to notify national competent authorities under the DGA) for interested practitioners and stakeholders.

Findings and conclusions

The findings of the report highlight the fragmentation and heterogeneity of the field. Data intermediaries that are currently most often addressed by practitioners and experts range from individualistic and business-oriented types to more collective and inclusive models that support greater engagement in data governance by communities and individual data subjects. Some data intermediaries focus the intermediation on technical solutions and infrastructures (e.g., PIMSs and data marketplaces), while others use legal constructs (e.g., data trusts), or other collective governance mechanisms (e.g., data cooperatives). Furthermore, while certain types aim at facilitating economic transactions between data holders and users, others mainly seek to produce collective benefits or public value.

In the report conclusions ([Section 5](#)), a series of **takeaways** are presented regarding the main obstacles faced by data intermediaries, and possible directions are identified for future science for policy research in this

field. The challenges faced by the entities addressed in the report include: **identifying suitable business models** that can guarantee economic sustainability; **extending the demand side** considering the widespread lack of awareness, interest, and expertise in data matters; and **comprehending the neutrality requirement** set by the DGA and how it could be implemented.

Finally, the report sets the stage for in-depth empirical research about data intermediaries, including collecting information from stakeholders potentially or directly involved in their use and development. Such research could assess the needs of developers and entrepreneurs for guidance on economic, legal and technical matters that play a role in establishing data intermediaries; the incentives for both the supply and demand side of data intermediaries; and the possible linkages of data intermediaries with other EU data policy instruments (such as the common European data spaces).

1 Introduction

This report examines emerging approaches to data governance based on different types of data intermediaries that, at least to some degree, may contribute to addressing the current imbalances of the data society and the economy. Data intermediaries can enhance access to data, improve control over how data is used, and therefore also increase both social and economic benefits generated by data, while assuming some of the costs and risks that come with such activities.

Data intermediaries are still relatively niche and embryonic initiatives, and therefore the knowledge and analysis about them is still evolving. On the one hand, in policy reports and academic literature, as well as among practitioners, various definitions and interpretations coexist about the different types of data intermediaries and how they function in theory. On the other hand, the empirical analysis of data intermediary case studies and the understanding of how they work in practice is still scarce.

Acknowledging the current status of the field, the key objective of the report is to contribute to establishing a common vocabulary, and to generate an understanding about the emerging landscape of data intermediaries among EU policy makers, experts, and practitioners. This work is part of the activities conducted by the JRC's Digital Economy Unit to support the implementation of the **Data Governance Act** (DGA) (European Commission, 2022). The Data Governance Act is a legislative initiative that was foreseen by the European Strategy for Data (European Commission, 2020b) and one of its main objectives is to increase trust in data sharing by fostering the establishment of 'data intermediaries' and

'recognised data altruism organisations'. It entered into force on 23 June 2022 and will be applicable from 24 September 2023. The timeline of this science for policy report is thus pivotal for informing the implementation of the DGA.

The report outlines the EU policy context and provides a landscape analysis of key types of data intermediaries, describing their main features and including relevant business model considerations. In order to inform future research activities, a series of thematic workshops and interviews will be conducted in the course of 2023 to gather further information and perspectives from data intermediary practitioners, e.g., to examine the impact of different business models on the sustainability of data intermediaries.

While recognising that multiple definitions and examples of data intermediaries exist, not all kinds of data intermediaries are addressed in this report. We have focused on those that are more widely discussed in the literature and that seem most closely aligned with the definition and objectives of the DGA. Six types of data intermediaries are presented: Personal Information Management Systems (PIMS), data cooperatives, data trusts, data unions, data marketplaces, and data sharing pools. The particular perspective adopted to examine the landscape of data intermediaries lies in the notion of inclusive data governance, which is explained in [Section 2](#) of the report.

The content presented in this report is the result of a combination of desk research and expert engagement. We are conscious of previous efforts by academics, policy officers, civil society actors, think tanks and foundations to analyse and classify emerg-

ing models for data governance and data stewardship: the report has been informed by their research and also benefits from the perspectives and inputs shared during an online expert workshop hosted on this topic by the JRC in March 2023.

This report wishes to position itself as a reference in this emerging field, not only by reviewing and condensing previous attempts to define data intermediaries, but also by adding key considerations about the business models of data intermediaries' and their overall economic sustainability, as well as by highlighting the relevance and contribution of current EU policy and legislative initiatives, in particular the Data Governance Act, to the understanding and possible evolution of data intermediaries.

Based on the literature reviewed for our mapping of **data intermediaries**, as well as the notion of inclusive data governance, in the report we adopt the following working definition of data intermediaries, which, although related, is broader than the definition contained in the Data Governance Act, acknowledging the variety of different models that currently exist in practice (see [Section 2.3](#)):

Data intermediaries for more inclusive data governance allow a broader range of stakeholders to access, control and share data, and support data subjects and data holders in deciding the purposes for which data is managed, as well as facilitating the exercise by data subjects of their rights over personal data, with the likely effect of producing further benefits from the same data and thus redistributing data value (social, public or private) across more actors and/or society.

We hope that this document can be useful for a wide array of stakeholders by helping them to navigate the landscape of data intermediaries. It aims to inform and raise

awareness about the potential and key features of data intermediaries, as well as about their challenges and limitations, and the ways in which the Data Governance Act could constitute a key step to support their development.

The report is structured as follows:

- [Section 1](#) introduces the report, describes the process through which it was developed and provides the policy context.
- [Section 2](#) clarifies the conceptual underpinnings of the report, explaining the perspectives adopted to examine the landscape of data intermediaries, which are anchored in the notion of inclusive data governance and in the Data Governance Act.
- [Section 3](#) defines the concept of business model, which is useful to analyse the strategies and conditions that could make data intermediaries sustainable.
- [Section 4](#) offers a review of current literature about data intermediaries, with the goal of identifying common elements and definitions agreed upon by scholars and policy makers. Six types of data intermediaries are presented: Personal Information Management Systems (PIMS); data co-operatives; data trusts; data unions; data marketplaces; and data sharing pools. For each type, the report provides general information about how it works and its main features, as well as selected examples and business model considerations.
- [Section 5](#) provides some concluding remarks, focusing on key obstacles for data intermediaries, and policy considerations and research recommendations for future work in this area.

1.1 EU policy context

Recognising data use as a key pillar for the European digital economy, the EU approach combines a push for the establishment of a single market for data (in which data can circulate freely and innovation can be fostered through access and use of data by more actors) with fundamental rights and freedoms to ensure a trustworthy and human-centric digital transformation.

One of the most influential regulations concerning the governance of (personal) data has been the EU General Data Protection Regulation (GDPR) (European Commission, 2016),¹ which entered into force in 2016 and has been applicable since 2018. The GDPR is designed to protect the data of EU citizens, creating a framework for enhancing trust by putting the interest of individual data subjects at the centre of the legislative text. In particular, Article 20 on the right to data portability has been a precursor to the establishment of data intermediaries for personal data within the EU context.

With regards to the topics addressed in this science for policy report, an important recent policy initiative is the [European Strategy for Data](#) (European Commission, 2020b), which foresees a key regulation that has shaped the thinking and practice around data intermediaries: the Data Governance Act. The European Strategy for Data was launched in 2020 with the aim of creating a single market for data that will ensure Europe's global competitiveness and data sovereignty. A single market for data and innovation can also be facilitated by SMEs, start-ups and European companies that develop and offer data intermediation services. Data intermediaries, in fact, support data access and the ability to (re)

reuse data, which is key for innovation and growth, and they have a great potential to strengthen the EU digital economy.

The following initiatives originated in the context of the European Strategy for Data and will impact on data sharing in Europe across actors, sectors and member states:

- The [Data Governance Act](#) (DGA)² was the first legislative act of the EU Data Strategy. It is a cross-sectoral instrument that aims at increasing trust in voluntary data sharing, for both economic growth and public interest purposes, by creating a framework for different data use and reuse situations. In particular, the Act outlines the following four goals:
 - making more data available by regulating the reuse of publicly held protected data;
 - boosting data sharing through the regulation of *data intermediaries*;
 - encouraging the sharing of data for altruistic purposes through the establishment of *recognised data altruism organisations*; and
 - facilitating data sharing across sectors and borders and enabling suitable targeted applications. Furthermore, the DGA will establish a European Data Innovation Board (EDIB) to facilitate the sharing of best practices (on topics covered by the DGA, like data intermediaries, data altruism and the re-use of public sector data) as well as on the prioritisation of cross-sectoral interoperability standards.

1. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX-3A02016R0679-20160504>

2. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex-3A32022R0868>

The Data Governance Act is the regulation that most explicitly addresses data intermediaries and thus can have a key role in promoting and sustaining them. The DGA proposal was published in 2020, it entered into force in June 2022 and **will be applicable from September 2023**. The DGA text has contributed to shaping the wider debate on the features that these new kinds of services for data intermediation should, or could, have. We will return to the regulation multiple times in this report, such as for the definition of data intermediaries ([Section 2.3](#)) and the understanding of the different types of data intermediaries ([Section 4](#)).

- The European Strategy for Data also foresees the establishment of **common European data spaces**, in which data intermediaries may play a key role in facilitating access to data. The Strategy sets out a vision for a ‘single European data space’, which it describes as ‘a genuine single market for data — open to data from across the world — where personal and non-personal data, including sensitive business data, are secure, and businesses have easy access to high-quality industrial data, boosting growth and creating value’. It emphasises that horizontal actions towards a single European data space should be accompanied by the development of sectoral or domain-specific data spaces in strategic areas - these are referred to as ‘common European data spaces’ (for a comprehensive overview and scientific insights on the common European data spaces, see Farrell et al., 2023). As described by the European Data Spaces Support Centre (2022), a common European

data space is a ‘decentralised, governed and standard-based structure to enable trustworthy data sharing between the data space participants on a voluntary basis’.

- Another key initiative within the European Strategy for Data is the proposal for the European **Data Act**.³ While still under negotiation at the time of writing, it contains measures designed to make the data economy more fair and innovative — for example, these measures include: allowing users to access data they have generated through the use of connected devices; rebalancing the negotiating power of SMEs by preventing the abuse of contractual imbalances in data sharing contracts entailing data access and use; allowing customers to switch between different cloud providers; and ensuring that the public sector can access privately held data in particular situations (for example, during public emergencies). The first of the above-mentioned four measures could be beneficial in the development of data intermediaries, as an increased access to data could mean greater power to decide who will be able to access it and how.
- Finally, another legal instrument arising from the EU Data Strategy is the recently published **Implementing Act on High Value Datasets**.⁴ The concept of high-value datasets was introduced by the Open Data Directive to describe datasets held by the public sector, the reuse of which holds

3. <https://eur-lex.europa.eu/legal-content/EN/TX/?uri=COM%3A2022%3A68%3AFIN>

4. European Commission, Commission Implementing Regulation (EU) 2023/138 of 21 December 2022 laying down a list of specific high-value datasets and the arrangements for their publication and re-use, Official Journal of the European Union, L 19, pp. 43–75 (https://eur-lex.europa.eu/eli/reg_impl/2023/138/oj) (accessed 1 February 2023).

the potential to generate important benefits for society, the economy and the environment. The availability of an increased number of datasets can represent an important resource for companies (in particular SMEs) to develop new digital products and services (including data intermediation services) and therefore it can be an enabler helping this new market to flourish.

- Two additional EU political agreements reached in 2022, although with different aims, will also impact on data and platform governance, and therefore on the development of data intermediaries. These are the **Digital Markets Act** (DMA)⁵ and the **Digital Services Act** (DSA),⁶ regulating market power and illegal content online respectively. The main goals of these two regulations are the creation of a safer digital space where the fundamental rights of all users of digital services are protected, as well as the establishment of a level playing field to foster innovation, growth, and competitiveness, both in the European Single Market and globally.

Beyond the specific legislative instruments briefly illustrated above, data intermediaries relate to and might play a role in achieving a diverse array of other digital policy goals:

- Firstly, the objectives and potential impacts of data intermediaries are related to the first priorities of the European Commission for the period 2019-2024⁷ **A Europe fit for the digital age** and **A European Green**

Deal. Fair and trustworthy data sharing through a new range of data intermediaries can not only support the development of innovative data driven products and services in Europe, but also unlock access to information and insights for addressing societal and environmental challenges. This can provide key benefits, ranging from personalised health care and improved mobility in urban settings, to better policy making. Data intermediaries could therefore play a role in fostering key objectives of the twin green and digital transition.

- Secondly, the agency of people *vis à vis* their data is mentioned in the **European declaration on digital rights and principles**⁸ signed by the European Parliament, the Council of the European Union and the European Commission on 15 December 2022. The Declaration promotes a digital transition shaped by European values and covers key rights and principles for the digital transformation, such as: *placing people and their rights at its centre; supporting solidarity and inclusion; ensuring the freedom of choice online; fostering participation in the digital public space; increasing safety, security and empowerment of individuals; and promoting the sustainability of the digital future.*⁹ With regards to data, the Declaration highlights the need to increase citizens' control over their data and its value, including by enhancing the use of data for public interest purposes, and recognising this as a form of empowerment.

5. https://ec.europa.eu/commission/presscorner/detail/en/ip_22_1978

6. https://ec.europa.eu/commission/presscorner/detail/en/ip_22_2545

7. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024_en

8. <https://digital-strategy.ec.europa.eu/en/library/european-declaration-digital-rights-and-principles>

9. https://ec.europa.eu/commission/presscorner/detail/en/IP_22_452

- Thirdly, **data sovereignty**, intended as enhancing control over data by organisations and individuals that contribute to their generation (Smith, 2016), implies addressing the asymmetries of power in the digital economy. This can be done for example through the regulation of gatekeepers¹⁰/Big Tech platforms; by allowing data subjects to have greater control over their personal data (GDPR); as well as by allowing data holders and users to engage in sovereign data sharing with others, through the promotion of data intermediaries and greater participation in data governance. The term, ‘data sovereignty’, is referred to in several EU policy initiatives (Madiega, 2020), such as: the European Strategy for Data¹¹, in its broader meaning of ‘digital sovereignty’, intended as a ‘means of promoting the notion of European leadership and strategic autonomy in the digital field’; as well as in the recent European declaration on digital rights and principles cited above (where it refers to the ability of Europe ‘to act independently in the digital world and should be understood in terms of both protective mechanisms and offensive tools to foster digital innovation’); and in the vision for the EU Digital Decade (which sets out a broader vision for EU digital transformation and establishes targets and actions to strengthen digital sovereignty). Finally, the President of the European Commission, Ursula von der Leyen, has mentioned digital and technological sovereignty in multiple

statements and contexts, describing it as ‘the capability that Europe must have to make its own choices, based on its own values, respecting its own rules’.¹²

It may be argued that the EU approach to data governance strives to find a balance between the protection of fundamental rights, e.g., privacy and data protection; security; safety, and ethical standards (Charter of EU Fundamental Rights¹³) and the promotion of the value brought by a more dynamic free flow of data. Following the analysis by Bodó et al. (2021) of the EU approach to data law, the European way may be described as a ‘mixed approach’ to regulating data and the digital transformation, which combines strong fundamental rights safeguards with a push for an EU internal market where data can circulate freely: the European approach to data governance) ‘envisions trustworthy data governance that reconciles responsible and human centric data governance, subject to full compliance with the EU’s strict data protection rules, while enabling data governance to foster innovation, and to drive economic growth (European Commission, 2020b)’ (Bodó et al. 2021, p. 2).

Beyond EU policy measures, an increased number of initiatives worldwide are advocating for the participation and empowerment of citizens in the governance of their data, both as individuals, and as members or representatives of groups and communities, and are raising awareness about how this might be achieved in different contexts and for different socio-demographics - this work can be supported by data intermediar-

10. Gatekeepers are digital platforms with a systemic role in the internal market that function as bottlenecks between businesses and consumers for important digital services (see the definition for the purpose for the Digital Markets Act, Article 3)

11. <https://digital-strategy.ec.europa.eu/en/policies/strategy-data>

12. Presentation of Commission’s strategies for data and Artificial Intelligence https://ec.europa.eu/commission/presscorner/detail/en/AC_20_260 State of the Union, 2021 https://state-of-the-union.ec.europa.eu/state-union-2021_en; State of the Union, 2020 https://state-of-the-union.ec.europa.eu/state-union-2020_en

13. https://www.europarl.europa.eu/charter/pdf/text_en.pdf

ies. For instance, scholars have advocated for children and their carers (parents and teachers) to be encouraged and engaged in controlling how their personal data are exploited by AI technology (Charisi et al., 2022), while civil society groups have advocated for local governments to adopt tools that ‘increase residents’ control over the personal data collected by the city and how it is shared’ (Cities Coalition for Digital Rights & UN HABITAT for a better urban future, 2022, p. 18).

1.2 What you will find in this report

The main goal of the report is to ‘set the scene’ and describe the main models and issues at stake in relation to the landscape of data intermediaries, in order to enhance a shared understanding and explore how they can promote more inclusive data governance. While certain topics covered in this report are already established among experts in the literature, it is also true that for many of them ambiguities and a lack of consensus still remain. An agreed terminology would ultimately be beneficial to people operating in such a relatively new field, whether researchers, practitioners or policy makers. This science for policy report thus wishes to raise awareness and increase the knowledge on current important issues relating to data governance.

The report includes a description of the main kinds of data intermediary models that, to a certain extent, can support more inclusive forms of data governance, such as those allowing increased control over data by a greater number of actors, and generating improved value production and distribution thanks to their re-use of data for different purposes. For each type of data intermediary, the report specifies the principal characteristics and key features (in-

tended as governance mechanisms and decision-making processes), while presenting selected examples and insights on business models and sustainability.

The process through which this report was created is the following:

- desk review and analysis of the main scholarly and policy literature by the JRC team;
- analysis of selected case studies from the information available online by the JRC team;
- feedback from EU policy officers working on data governance;
- feedback from experts, researchers and academics working on data governance and related topics.

The report sets the stage for a more in-depth empirical analysis of these data intermediaries, as well as of the stakeholders involved in their use and development. Future activities for this project will be based on information directly collected from data intermediary practitioners, developers, managers and users about their first-hand experiences and perspectives, which will then also be analysed through the lens of the issues detailed in this document.

2 Towards more inclusive data governance

2.1 Data governance

While there may be different definitions, we understand data governance as comprising the set of rules, policies, legal mechanisms, stakeholders relations, decision-making structures and processes established to collect, share and use data. It concerns how data can be managed by data subjects and holders, as well as how data may be made accessible to third parties. Data governance can be implemented through different means, which range from contractual agreements to technical standards, from consent forms to goals and principles. Ultimately, data governance determines the value produced from data and how it is distributed between actors and across society.

We understand data governance not only as the regulations at the macro-level (i.e., the set of rules of a legislative nature established by policy makers), but also as the relations, frameworks, principles and socio-technical arrangements established by multiple social actors in order to manage data. This understanding is informed by the notion of governance in political science, which acknowledges that a broad set of actors and institutions, beyond regulators, including the private sector, civil society and other non-governmental entities, are involved in managing societies (Kooiman, 2003).

Until a decade ago, data governance was mainly understood at the micro-organisa-

tional level (especially in the fields of information systems and privacy regimes) as decision rights and accountabilities for managing data as a strategic asset of a single enterprise (Abraham et al., 2019; Zygmuntowski et al., 2021). Data governance was mainly seen as the internal task of a single company or organisation, which has to control and manage data in an efficient way (Data Administration Management Association, 2009). Lately, the concept has also been applied in a wider context at the meso-level, encompassing the social, economic, political, and cultural factors that explain how data are accessed and controlled across organisations and by a multitude of actors, beyond the single enterprise (Liu, 2022). In this report, we adopt this latter understanding, according to which data governance concerns trans-organisational data flows at different levels examined through a holistic approach that also accounts for the influence of the micro and macro levels (Zygmuntowski et al., 2021; Lähteenoja and Sepp, 2021).

The notion of data governance recently gained popularity because of the growing recognition of the exploitative and monopolistic practices that characterise our digital ecosystem and the urgency to explore alternative ways of controlling data and creating value from them (Ada Lovelace Institute, 2022; Micheli et al., 2020; Zygmuntowski et al., 2021). Various data governance models have been identified that allow stakehold-

ers to establish more equitable and participatory relations compared to those promoted by Big Tech corporations, which tend to be highly asymmetric and monopolistic. These new data governance approaches range from horizontal collaborations between businesses and public entities, to data stewardship for the responsible management of data on behalf of data subjects, 'pioneering attempts at creating collaborative governance regimes with public interest in mind' (Zygmuntowski et al., 2021, p. 7). Data intermediaries are expected to play a role in this field and a heterogeneous landscape of data intermediation initiatives has already been launched with similar goals in mind. In this report we explore different types of data intermediaries that promote more horizontal stakeholder relationships and fairer ways to handle data, thus contributing to advancing more inclusive forms of data governance.



2.2 Inclusive data governance

The review of data intermediaries presented in this report is informed by the notion of inclusive data governance. With this concept, we refer to initiatives for data sharing, control and use that promote alternatives to dominant monopolistic and exploitative data practices. Data governance approaches can be considered more inclusive for different reasons: because they support more horizontal relations between actors involved in data collection and use; because they offer means to access and share data assets for smaller economic players in the data market; because they foster data sharing for public interest purposes; because they allow data subjects to decide about how to use their data; and because they support communal approaches to data management and sharing based on transparency and accountability, etc.

Data governance is more inclusive when there is a broader stakeholder participation in decisions concerning data access, control, sharing and use, as compared to the 'take it or leave it' approach to data typical of large technology platforms. This could imply a more horizontal relationship between suppliers (data holders and data subjects), as well as increased opportunities and means to access and control data for data subjects and other actors. A more inclusive data governance approach allows multiple stakeholders to use data for various purposes. Consequently, it might result in the generation of different kinds of economic and social value, which could be understood as a 'redistribution' of the benefits generated from data. This is an important step for a fairer data ecosystem in which data are managed for the benefit of all, not the few.

The notion of 'inclusive data governance' is used in this report as a conceptual lens. It is intended to describe data handling practices, aligned with European principles and values, that promote a fairer data economy and society. We consider inclusive data governance as a potential paradigm shift that could be achieved through different kinds of initiatives. Data intermediaries play a role in fostering inclusive data governance, but this shift can also be supported by wider legal and policy interventions, such as regulation of unfair data practices, online content regulation, or policies ensuring that the profit that data help to generate is redistributed more equitably. Data intermediaries cannot, by themselves, undermine the exploitative practices and power symmetries of the current digital economy (Janssen et al., 2020). However, they support more inclusive approaches for handling data, which are alternative to the dominant ones. If data governance in general implies issues of data control and value generation, inclusive data governance means configuring them ac-

ording to specific ways and motivations. We elaborate more on two positive features of inclusive data governance below:

● Control and agency:

- Data governance approaches are inclusive when a broader range, and more diverse types, of actors can access data and decide how data are shared, accessed and used, compared to unilateral agreements typically set by Big Tech platforms.
- Data governance is inclusive when the interests, needs and rights of those who currently have less power in the data economy (such as citizens, civil society organisations, but also SMEs and local public authorities) are protected and promoted.
- Data governance is inclusive when data holders, data subjects, and all individuals represented in data, or affected by how data are used, can rely on governance mechanisms and tools that allow them to have agency over their data.
- For individuals and communities, data governance is inclusive when there is participation around data. Participation takes place in different forms, -from public dialogues and ethics committees to clear terms of services and privacy-enhancing technologies, and from individuals controlling their personal data to collective forms of bargaining over data rights and collaborative decision-making on data commons.

- In the economic context, data governance is inclusive when SMEs and start-ups have greater control over their data, and when they can access relevant data sources and use them to build functional (and socially relevant) data-driven services instead of being subjected to the information asymmetries of data monopolies.
- Greater agency and control over data by both individuals and (smaller) economic actors could enhance the establishment of a healthier data ecosystem that is 'fundamental rights compliant', in which innovation is fostered and in which some of the inequities in how data are used, collected and shared are addressed.

● Value and benefit sharing:

- Data governance approaches are inclusive when different kinds of value are generated from data across multiple value chains and sectors of society. The different types of value range from economic and social to moral, including the ability to negotiate privacy.
- Data governance is inclusive when the value generated from data is spread across a wide range of actors and collective claims can be made about value creation and distribution.
- Data governance is inclusive when the outcomes of data use are distributed fairly, for instance when some of the profits that emerge from commercial data are returned to the public

domain or to other actors that enabled the production of these data in the first place (e.g., from public infrastructures to citizens' digital footprints). This resonates with notions such as 'data solidarity' and 'data as utility'.

- Data governance is inclusive when economic value generated from data use is created through partnerships established to optimise data value chains in a sector or logistics.
- In the context of this report, inclusive data governance refers both to 'economic inclusion' (i.e., increased participation in value creation by data subjects and data holders thanks to greater opportunities to obtain value from data), as well as to public value creation. Data sharing for the benefit of citizens and society encompasses harnessing data for the common good, beyond solely economic benefits, through altruistic forms of data sharing to address societal challenges (in sectors such as health and environment).

Both features described above (Control and agency, as well as Value and benefit sharing) align with current EU data policies, which are based on a human-centric approach to the digital transformation and anchored in European values and fundamental rights. Putting individuals at the centre of (data) governance resonates with this broad EU vision for a digital economy and society, and the empowerment of both individuals and businesses regarding the data they contribute to producing is an important

theme in the European Strategy for Data.¹⁴

In this context, the rights attributed to data subjects for the control over their personal data by the GDPR have been an important key step. In particular, the rights to data access and to data portability have been established with the goal of empowering data subjects to control their data by enabling them to access -and then move, copy or transmit personal data easily from one digital environment to another (Giannopoulou et al., 2022). The Data Governance Act promotes those principles and governance features described above via the concept of trustworthy data intermediaries. The DGA aims to push forward data intermediation services that are 'fundamental rights compliant' as they respect human dignity and personal autonomy- values that lie at the heart of the EU Charter of Fundamental Rights.

Data intermediaries offer technical and legal means to control and decide about data, opening up new channels for data sharing and for distributing the benefits generated by data more widely across the economy and society. One of the intentions of the DGA is to provide the means to enhance trust in these new data services by legally separating (neutral) data intermediaries from other data governance models that are based on power or information asymmetries. The DGA could thus potentially contribute to supporting the uptake of more inclusive data governance approaches in the data economy and society.

14. Both features are 'positive' aspects of data governance (who has the power and who gets the benefits). Although we do not delve into this aspect, we recognise that it might be valuable to also examine the topic from the perspective of 'negative' features, namely who carries the obligations and who bears the risks and the costs. These 'negative' features shall also be more equitably distributed in an inclusive model of data governance.

In addition to regulatory measures, technical means may also foster inclusive data governance. Although technical aspects will not be addressed in detail in this report, we should stress that the adoption of application programming interfaces (APIs) and interoperability standards could play a fundamental role in promoting inclusive data governance, preventing lock-in effects and supporting fairer competition, and enabling data sharing among different data holders and data users (Borgogno and Colangelo, 2019).

2.3 Data intermediaries

Data intermediaries are a relatively new kind of entity in the data economy for which multiple definitions exist. In this section, we first outline how data intermediaries are understood and regulated in recent EU legislation, then we draw from conceptualisations currently widespread in the academic and policy literature to highlight how this new kind of service might promote more inclusive approaches to the governance of data, and finally we conclude by proposing an operational definition of intermediaries based on the above.

In the first section (2.3.1) we tackle data intermediaries from the perspective of the Data Governance Act. In addition to summarising the objectives of the regulation, the section describes the requirements that data intermediaries must abide by according to this legislation. Only data intermediaries that meet certain criteria are covered by the DGA. We refer to this subset of intermediaries as **data intermediation services providers** (DISPs), as this is the terminology used in the DGA text.

The second section (2.3.2) reviews data intermediaries as defined in the academic and policy literature, using the lens of what

we call '**inclusive data governance**' (see [previous section](#) for a definition). It provides an overview of the different roles that data intermediaries could play in relation to redressing the information and power asymmetries that characterise today's political economy of data. This section identifies four objectives that can be pursued by data intermediaries that foster more inclusive approaches to data governance. It then closes by providing an operational definition of data intermediaries adopted in the current report. Importantly, intermediaries covered by the DGA regulation can also be intermediaries promoting inclusive data governance; the distinction is analytical, but the two concepts could apply to the same initiative in practice (see [Figure 1](#)).

2.3.1 Data intermediaries and data altruism organisations in the Data Governance Act

The Data Governance Act (DGA) aims to stimulate the availability of data for (re)use and to strengthen data governance mechanisms in the EU. It is designed to facilitate data sharing by/between different actors (businesses, individuals, public sector entities, NGOs) by enabling more trustworthy data exchanges for different purposes. The legislation is part of the European Strategy for Data, thus it is based on the premise that data could lead to beneficial societal impacts — such as improvements in health and well-being, strengthened climate action and more efficient public services, as well as fostering innovation and economic growth, with data being a critical resource for start-ups and SMEs. Yet, as stated in the impact assessment of the DGA, the potential of data is not fully released in the EU due to several obstacles to data sharing (European Commission, 2020a).



Low trust in data sharing has been identified as one of the main reasons behind limited data availability and use in Europe (European Commission, 2020a). The DGA puts forward a variety of measures with the explicit intention of promoting increased trust in data sharing among businesses, organisations and individuals. Concerning data intermediaries, it defines specific rules for ‘data intermediation services providers’ (**DISPs**) in Chapter III (see [Box 1](#)). Chapter IV of the regulation addresses instead entities that enable data altruism, which will be able to register as ‘data altruism organisations recognised in the Union’ (**RDAOs**) (see [Box 2](#)). The regulation makes a clear distinction between DISPs and RDAOs according to their objectives and functionalities. It defines a ‘data intermediation service’ as a service which aims to establish ‘commercial relationships’ for the purposes of data sharing between data subjects/holders and data users, while ‘data altruism’ is defined as the sharing of data without seeking a reward, ‘for objectives of general interest’.

DISPs and RDAOs are legally binding definitions (for services that are covered by the regulation in the first case, or wish to register as such in the latter) and apply to a specific set of data services, belonging to — but not fully overlapping with — the wider landscape of data intermediaries that is reviewed in this report, which focuses on inclusive data governance. In this section, we clarify which entities are recognised under the DGA, how they relate to each other, and how they relate to a wider landscape of data intermediaries.

2.3.1.1 Data Intermediation Service Providers (DISPs) regulated by the DGA

The DGA advances a governance framework that ‘offers an alternative model to

the data-handling practices of the Big Tech platforms’¹⁵ through the promotion of neutral and trustworthy data intermediation services providers that put individuals and companies in control of their data. One of the main objectives of the DGA is to identify a new category of neutral data intermediaries that comply with a specific set of rules and are thus recognised by the EU (for a summary see [Box 1](#)). Companies, organisations and individuals will thus have the possibility to rely on such **neutral and trustworthy intermediation services** for the sharing or pooling of data as well as for exercising their rights as data subjects. This new framework is developed to reassure data subjects/holders about how their data will be treated — i.e., they will know that they interact with an intermediary that is managing their data according to high security standards, does not have any conflict of interests, and that fully complies with the applicable EU regulations.

In practice, the DGA requires that **data intermediaries** operating in the EU which aim to establish **commercial relationships** for the purposes of data sharing between data subjects/holders and data users must undergo a **mandatory notification procedure** to the competent authority. The provider of the data intermediation services must submit a notification containing a specific set of information about those services, such as legal status, address of main establishment in the Union, a designated contact person, and a description of the type of services provided. Once the authority confirms that the service complies with Articles 11 and 12 of the DGA, the DISP will be able to use the label ‘data intermediation services provider recognised in the Union’ and to display the related logo. The Com-

15. <https://digital-strategy.ec.europa.eu/en/policies/data-governance-act-explained>

mission will keep a regularly updated public register of recognised DISPs, which will allow data holders and subjects to find trustworthy operators of data intermediation services.

Registered DISPs should belong to one of the three types of data intermediaries described in Article 10 of the DGA (see [Table 1](#)) and should also comply with the **conditions** listed in Article 12 of the DGA, which are set out to ensure that there are no conflicts of interest.

A key condition for DISPs is that they shall be **'neutral third parties'**, which means they are only operating as intermediaries, without processing any of the data themselves to obtain data-driven services or products from the information provided. DISPs cannot monetise data by selling it to another company or by using it to develop their own data-driven products or additional services based on this data. DISPs shall not intervene in the data value chain unless these additional data-driven services are operated through separate legal entities: i.e., there must be a **structural separation** between the data intermediation services and any other data services provided. For companies that already offer data intermediation services in addition to their other data services, 'the data intermediation activity must be strictly separated, both legally and economically, from other data services'.¹⁶ A key objective of these DGA requirements is

16. <https://digital-strategy.ec.europa.eu/en/policies/data-governance-act-explained>. Additional conditions summarised at this link include that the commercial terms (such as pricing) for the provision of intermediation services should not be dependent on whether a potential data holder or data user is using other services, and that any data and metadata acquired by DISPs can be used only to improve the data intermediation service itself.

for data holders/subjects to know that the business model of a DISP does not depend on making a direct profit from the information it is sharing. The only beneficiaries of the data value shall be the data suppliers and selected/known data users, not the intermediation services providers that enable the data sharing.

This set of conditions clearly distinguishes DISPs from those large online platforms that qualify as gatekeepers according to the Digital Markets Act, as well as from other kinds of data intermediaries that do not meet the DGA criteria. For instance, data brokers, which are widely acknowledged as data intermediaries in the academic literature (Janssen & Singh, 2022a; von Ditfurth and Lienemann, 2022), might fall outside the DGA mandate in case their goal is not to establish a commercial relationship between data subjects/holders and data users, or whether they aggregate, enrich or transform data with aim of adding value and licencing its use to data users (see Art 2(11)a). The types of data intermediaries falling under Chapter III of the DGA represent only a segment of the wide spectrum of data intermediaries operating in the data economy. It is important to highlight that, although DISPs cannot implement the business model of the data brokers mentioned above (or generate a profit from other data services), they still need to find a revenue stream to be economically sustainable as private companies (for a general overview of business models of different types of data intermediaries and the challenges for reaching economic sustainability see [Section 3](#)).

Box 1. The Data Governance Act's provisions on data intermediation services providers (DISPs) in brief

Article 2 of the DGA defines a data intermediation service as

a service which aims to establish commercial relationships for the purposes of data sharing between an undetermined number of data subjects and data holders on the one hand and data users on the other, through technical, legal or other means, including for the purpose of exercising the rights of data subjects in relation to personal data (Art. 2(11)).

From this provision it follows that an essential criterion for an entity to be defined as a data intermediary according to the DGA is the establishment of commercial relationships between data subjects/holders on one side and data users on the other for the purpose of intermediation.

In Article 10, the DGA foresees three broad types of data intermediation services, which can be conceptualised as 'enablers of data spaces' following the reasoning described in the Recitals of the DGA, these are:

- Intermediation services between **data holders** and potential data users; this concerns entities that allow bilateral or multilateral exchanges of data, the creation of platforms or databases enabling the exchange or joint use of data, as well as the establishment of other specific infrastructure for the interconnection of data holders with data users.
- Intermediation services between **data subjects or individuals** that seek to make their personal or non-personal data available, and potential data users.
- Data cooperatives, which are organisational structures constituted by data subjects, one-person undertakings or SMEs. These entities will help **members of the cooperative** to exercise their rights over their data.

While the first type can be enablers for industrial data sharing, the second mainly focuses on personal data sharing, and the third covers collective data sharing and governance schemes.

The Recitals of the DGA also mention specific types of data intermediaries, such as: data marketplaces, orchestrators of data sharing ecosystems (data spaces) and data pools (Recital 28), which could be examples of the first type, and personal data spaces (Recital 30), which could be an example of the second type.

Article 12 of the DGA specifies that data intermediaries can only be mediators (*neutral third parties*) and cannot *aggregate, enrich or transform the data for the purpose of adding substantial value to it and license the use of the resulting data to data users*.¹⁷

If an entity complies with definitions in Art. 2(11) and Art. 10 it is expected to follow conditions in Art. 12 to be able to register with the competent national authority for data intermediation services under the DGA. This requirement applies to entities providing data intermediation services in the EU, regardless of where they are established.

17. <https://digital-strategy.ec.europa.eu/en/policies/data-governance-act-explained>

2.3.1.2 Data altruism organisations recognised in the EU (RDAOs) by the DGA

The DGA also advances a governance framework to create trusted tools for data altruism, intended as the voluntary sharing of data without reward for objectives of general interest. The notion of ‘**general interest objectives**’ is left for national law to define, but, overall, such objectives can include healthcare, fighting climate change, augmenting official statistics, improving mobility and public services, as well as enhancing public policy making and research. Chapter IV is dedicated to organisations establishing data altruism initiatives of various kinds, which will have the possibility to register as ‘data altruism organisations recognised in the Union’ (RDAOs) (for a summary see [Box 2](#)).

Unlike DISPs, RDAOs must be run by a not-for-profit entity and their registration under the DGA is **voluntary**. Only organisations that meet a set of **transparency requirements** (Art. 20), that offer specific **safeguards** to protect the rights and interests of individuals and companies who share their data (Art. 21), and that comply with the Data Altruism **Rulebook**¹⁸ (Art. 22 and delegated act) will be able to register as RDAOs. The Rulebook details technical, security and information requirements, together with communication roadmaps and interoperability standards. Once an entity is registered as an RDAO, it will be able to use a special logo and will be listed in both the Commission register and in the Member State public national register for RDAOs. These registers support data holders/subjects in finding trustworthy tools for data sharing for the public interest. A **European data altruism consent form** (Art. 25) will also be adopted to allow the collection

of consent or permission across Member States in a uniform format. Overall, DGA Chapter IV measures are meant to increase trust among data subjects/holders towards data altruism through the identification of recognised data altruism organisations.

For the objectives of this report, it is useful to clarify how RDAOs relate to DISPs and data intermediaries more generally. One of the key requirements for qualification as a recognised data altruism organisation is that an entity must operate on a **not-for-profit basis** and be legally independent of any entity that operates on a for-profit basis. This neatly demarcates RDAOs (data altruism organisations covered in Chapter IV) from DISPs (data intermediation services providers covered in Chapter III), but not from data intermediaries more generally. It is also a requirement that data altruism activities are carried out through a structure that is separated from other activities, to avoid conflicts of interest and help increase trust.

RDAOs can operate in different ways, with a broader range of roles than being a data intermediary. According to details in the legislation (see for instance Articles 20 and 21), RDAOs can perform the following activities (while ensuring they comply with the requirements outlined above):

- *collect* and *store* data from data holders/subjects for objectives of general interest;
- *process* data collected from data holders/subjects for objectives of general interest;
- *allow* third parties (natural or legal persons) to process data collected from data holders/subjects for objectives of general interest.

Only the last of the activities listed above implies data intermediation. We thus conclude that a RDAO might, in certain cases, carry out general data intermediation activ-

18. Compliance with the delegated act (Rulebook) referred to in Article 22 must take place at least 18 months after its entry into force.

ities (broadly speaking, we are not referring here to data intermediation services regulated by Chapter III of the DGA) in the public interest, but it is not necessary to do so in order to qualify as RDAOs. Furthermore, RDAOs are clearly distinguished from data intermediation services providers (DISPs) that facilitate the establishment of com-

mercial relationships between data suppliers and users. RDAOs can both process and use data, if they do so according to the requirements set out in the regulation, while DISPs can process data but, in principle, they cannot use it (or only to the limited extent to improve their services).

Box 2. The Data Governance Act's provisions on data altruism organisations (RDAOs)

Chapter IV of the DGA provides a framework for increasing trust in data altruism organisations. The regulation establishes a voluntary labelling scheme to which entities that carry out data altruism initiatives can apply so that they can be registered as 'Data altruism organisation recognised in the Union'. The objective of this labelling framework is to allow individuals and companies to share their data for the public good with increased transparency and trust.

The key feature of data altruism is a focus on the common good as opposed to seeking economic benefits for the individual stakeholders engaged in data sharing. Data altruism is defined in **Article 2** as *the voluntary sharing of data on the basis of the consent of data subjects or permissions of data holders to allow the use of their non-personal data without seeking or receiving a reward and for objectives of general interest as provided for in national law, where applicable, such as healthcare, combating climate change, improving mobility, facilitating the development, production and dissemination of official statistics, improving the provision of public services, public policy making or scientific research purposes in the general interest.*

Entities that make data available through data altruism and that comply with the requirements set out in the dedicated Rulebook (**Article 22**) will be able (but not obliged) to register as 'Data altruism organisation recognised in the Union (RDAO)' with the national competent authority for the registration of data altruism organisations. The registration is not a compulsory provision, and therefore being registered is not a requirement to conduct data altruism initiatives. The registered organisations, however, will ensure a higher level of trust for data subjects and data holders compared to non-registered ones.

As defined in **Article 25**, data altruism organisations recognised in the EU will provide adequate means to data holders and data subjects to withdraw or modify their consent, including updates about the use of their data. The DGA foresees a European data altruism consent form (which will be adopted by the European Commission through an implementing act) to smooth the collection of consent (for personal data) and permission (for non-personal data). The consent form is also intended to facilitate the portability of data in order to allow data subjects to leverage Article 20 of the GDPR. With more control over their data, data holders and data subjects are expected to be more likely to trust other agents handling them, and thus to share their data for the public good, since they can reverse or amend their actions if expectations are not met.

Table 1. DGA provisions for DISPs and RDAOs

	Providers	Data suppliers	Purposes for data sharing
DISPs <i>Data intermediation services providers recognised in the Union</i>	<ul style="list-style-type: none"> Providers of a service which aims to establish commercial relationships for the purposes of data sharing between data subjects/holders and data users through technical or legal means. Neutral third parties (structural separation between data intermediation services and any other data-driven services). <p>DISPs do not include:</p> <ul style="list-style-type: none"> services that aggregate, enrich or transform data for adding value; services that focus on the intermediation of copyright-protected content; services that are exclusively accessed by one data holder; data-sharing services offered by public sector bodies that do not aim to establish commercial relationships; data altruism organisations. 	Data holders	<ul style="list-style-type: none"> Bilateral or multilateral exchanges of data; creation of platforms or databases enabling the exchange or joint use of data; establishment of other infrastructure for the interconnection of data holders with data users.
		Data subjects and individuals	Make personal or non-personal data available for potential data users.
		Groups of data holders or data subjects ¹⁹ (Members of data cooperatives)	<p>Support members in the exercise of their rights over their data, with regard to:</p> <ul style="list-style-type: none"> making informed choices before they consent to data processing; exchanging views on data processing purposes and conditions that best represent the interests of its members; negotiating terms and conditions for data processing on behalf of its members.
RDAOs <i>Data altruism organisations recognised in the Union</i>	<ul style="list-style-type: none"> Entities operating on a not-for-profit basis enabling the voluntary sharing of data for objectives of general interest (e.g., healthcare, fighting climate change, official statistics, improving mobility and public services, enhancing public policy making and research). Complying with Rulebook, meeting transparency requirements, and offering specific safeguards. 	Data holders and data subjects	Consent or permission to collect, store, process or share data with third parties for objectives of general interest.

Source: JRC own elaboration with contribution of Viivi Lähteenoja

19. Data subjects, one-person undertakings or SMEs.

2.3.2 Data intermediaries for more inclusive data governance

This report provides an overview of data intermediaries that, to different extents, foster alternatives to the exploitative asymmetric data practices of corporations and very large online platforms (Liu, 2022; Craglia et al., 2021). In recent years, several authors and organisations made attempts at systematising knowledge on the topic of data intermediaries, elaborating their own definitions and perspectives on the role that these new types of services could play in fostering different and more inclusive approaches for the governance of data. This section reviews some of the most common conceptualisations of the notion of data intermediary informed by that perspective. At the end of the section, a short definition, based on the academic and grey literature previously reviewed, is proposed to summarise how data intermediaries for more inclusive data governance are conceived in this report. Importantly, these data intermediaries make up a different group than the services (DISPs) regulated by the Data Governance Act, but the two groups overlap (see [Figure 1](#)).

In a recent article titled ‘Data Intermediary’ published in the journal ‘Internet Policy Review’, Janssen and Singh (2022a) provide the following definition of data intermediaries, stressing the roles of the different actors involved in the data flows:

A data intermediary serves as a mediator between those who wish to make their data available, and those who seek to leverage that data. The intermediary works to govern the data in specific ways, and provides some degree of confidence regarding how the data will be used. (...) Data intermediaries form part of a data processing ecosystem. This includes the intermediary, often an

organisation (of some form), as well as two other key categories of stakeholder: data suppliers who are those individuals, communities, or enterprises that make their data available, and third parties referring to those interested in using (processing) supplier data. (Janssen & Singh, 2022a, p. 2).

Data intermediaries might foster more inclusive data governance in different ways. While reviewing some of the key definitions, we identified the following objectives that also support inclusive data governance goals:

- A key objective of data intermediaries is to enable **data subjects** and individuals to exercise **greater control over their data**. A data intermediary is a trusted mediator that enables individuals to take decisions about their data, such as about how these are accessed, with whom they are shared, how they are used and for which purposes, facilitating consent and decision-making, while preserving the right to data privacy. The data portability right (enshrined within the European Union by the General Data Protection Regulation - GDPR²⁰) sets the conditions for such data intermediaries to exist, as it provides data subjects with the right to access their data, aggregate them and use them for other purposes.²¹ This kind of data intermediary might include different types that operate at the individual or collective level, such as Personal Information Management Systems

20. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX-3A02016R0679-20160504>

21. Article 20 of the GDPR states that a data subject has the right to obtain personal data concerning him/her from a controller and to store or transmit them to another controller. In this way it allows individuals to obtain (in a machine-readable way) and reuse their personal data for their own purposes across different services.

(PIMS) (and services labelled in a similar way, such as personal data wallets or clouds), data unions and data cooperatives.

- Other definitions highlight the impact that this new type of actor might have on the political economy of the current data ecosystem by addressing information and **power asymmetries**: data intermediaries can ‘help rebalance relationships between those producing or with rights over data, and those seeking to use that data by offering an alternative approach to the data processing’ (Janssen and Singh, 2022a; see also Mozilla Insights, van Geuns and Brandusescu, 2020). To promote a more balanced ecosystem, data intermediaries should not have conflicts of interests over the data they intermediate. A recent report by the Ada Lovelace Institute (2022, p. 58) stated that data intermediaries have to be neutral, i.e., ‘discouraged and prevented from monetising data’. If intermediaries do not derive profit from data aggregation, transformation and use, they will be clearly distinct from the gatekeeper platforms that dominate the market.
- Data intermediaries, by facilitating sharing, also enable the potential for data re-use and for the **generation of additional (kinds of) value** from data for both the economy and society. A typical example of data intermediation services that facilitate the creation of more economic benefits are those in the business-to-business context, such as data marketplaces or data pools, which allow the establishment of relationships between multiple data holders and data users, including smaller economic players. A definition of data intermediaries that

places more emphasis on value creation is proposed in the OECD (2019) report entitled *Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies*:

Data intermediaries enable data holders to share their data, so it can be re-used by potential data users. They may also provide additional added-value services such as data processing services, payment and clearing services and legal services, including the provision of standard license schemes (OECD, 2019, P. 36).

Besides generating economic benefits to companies, data intermediaries might also produce societal benefits. By facilitating data access and sharing, they create opportunities for better data analysis in different areas, including research, policy decision-making and technology audit. A definition provided by the Centre for Data Ethics and Innovation of the UK Government (2021), for instance, acknowledges how data intermediaries offer technical, legal and economic means for improved data analysis as value creation.

Intermediaries can provide technical infrastructure and expertise to support interoperability between datasets, or act as a mediator negotiating sharing arrangements between parties looking to share, access, or pool data. They can also provide rights-preserving services — for example, by acting as a data custodian allowing remote analysis through privacy-enhancing technologies, or providing independent analytical services in a siloed environment (Centre for Data Ethics and Innovation, 2021)

- Data intermediaries support the **responsible management** of information resources and are also referred to as data stewardship institutions or collectives (Hardinges, & Keller, 2021). Given the role they can play in protecting the (data) rights of individuals and communities and in unlocking the benefits produced by data, they are expected to foster more trustworthy, responsible, fair and equitable data use. Data stewardship ‘denotes an approach to the management of data [...] intended to convey a fiduciary (or trust) level of responsibility toward the data’ (Rosenbaum, 2010 cited in UK AI Council & Ada Lovelace Institute, 2021, p. 19). Data stewardship, however, is not just as an institution, but also a practice that facilitates establishing a responsible culture for data access and re-use. It is also referred to as a professional practice with certain roles, skills and responsibilities (Verhulst, 2021).

Data intermediaries allow the establishment of a relationship (commercial or non-commercial) between data subjects and/or data holders, on the one hand, and data users on the other hand. Data intermediaries for more inclusive data governance allow a broader range of stakeholders to access, control and share data, and support data subjects and data holders in deciding the purposes for which data is managed, as well as facilitating the exercise by data subjects of their rights over personal data, with the likely effect of producing further benefits from the same data and thus redistributing data value (social, public or private) across more actors and/or society.

Data intermediaries are a wide category of services. This report maps the broader landscape of these services that support more inclusive forms of data governance and thus promote a human-centred and ‘fundamental rights compliant’ data society, in which data is controlled and leveraged by individuals, communities, SMEs and other actors for the generation of social benefits, while also innovating in the digital economy. These might overlap with DISPs covered by the DGA, depending on the extent to which they comply with its requirements. The relationship between the two groups is depicted in [Figure 1](#).

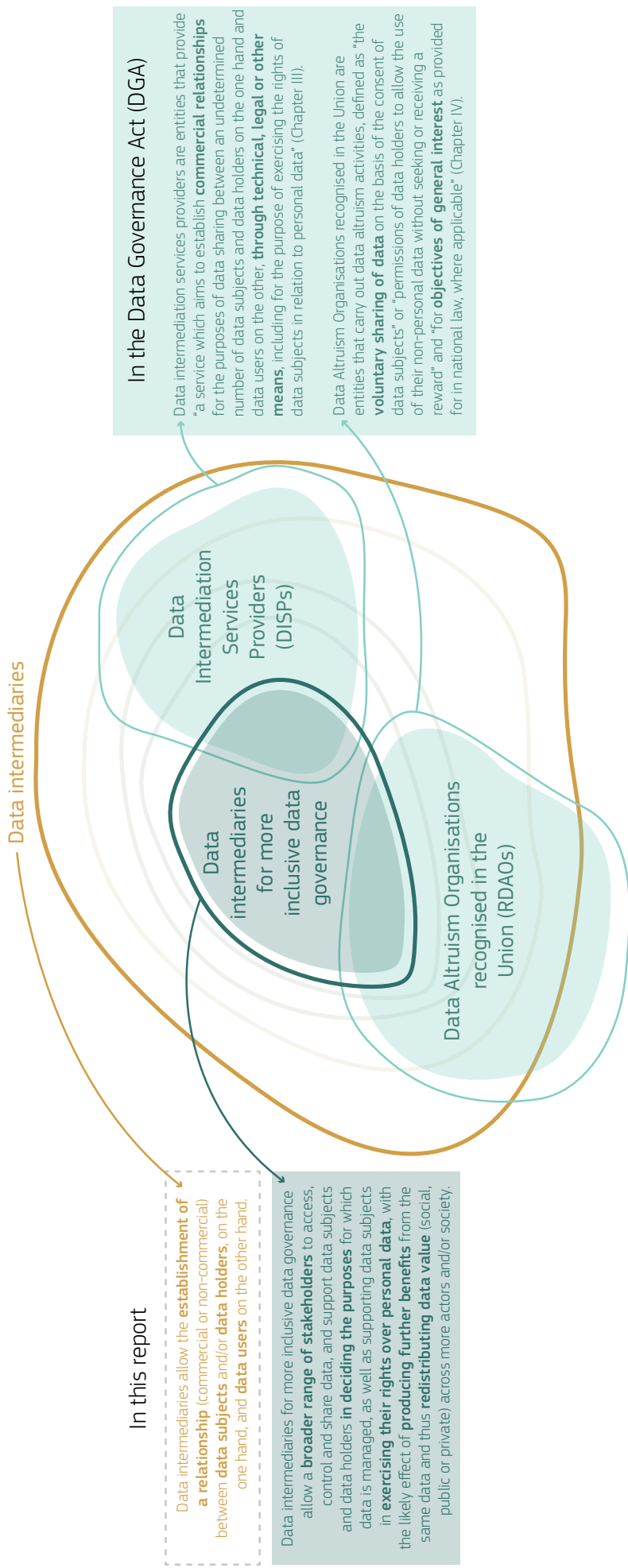
It is worth noticing that pursuing one of the above objectives does not rule out the others. On the contrary, several types of data intermediaries are designed to address most, if not all, the above-mentioned goals. For example, Giannopoulou et al. (2022) explain how data trusts (Delacroix and Lawrence, 2019) may support the ‘greater control over their data’ objective through rights mandates while also ‘addressing power asymmetries’ and supporting the ‘responsible management’ of information.

Drawing from the strand of literature briefly summarised above, this report refers to data intermediaries for more inclusive data according to the following definition:



Figure 1

Overlaps between the different definitions of data intermediaries in this report.



Source: JRC own elaboration

3 The economic sustainability of data intermediaries

In this section we study the challenges and opportunities that data intermediaries face to reach economic sustainability. Our focus is on data intermediaries in general, which includes but is not limited to the data intermediaries considered in Chapter III of the DGA, as shown in [Figure 1](#).

Reaching economic sustainability is one of the main challenges that data intermediaries currently face in order to become relevant actors and hence deliver the expected positive impacts referred to above. In this section we build on the concept of business model in order to analyse the strategies and conditions that could make data intermediaries economically sustainable. We begin by defining a business model and its components, which we exemplify for the case of a data intermediary. We then describe each of the core components of a business model in detail with a particular focus on the challenges that data intermediaries have to face in order to develop sustainable business models.

Regardless of what the main objective of a data intermediary is, it needs to design an economically-sustainable business model if it wants to survive. This has important implications for their business model design when they aim at being enablers for inclusive data governance. First, because several of these types of data intermediaries have a major non-economic value focus. Second, given that data intermediaries

aim at giving control to several stakeholders through a neutral organisation, business practices that benefit a single or a few actors to the detriment of other stakeholders are not compatible with the notion of inclusive data governance. Third, as detailed in [Section 3.2.1](#), the nature of data intermediary activities structures their possible value proposition. This, in turn, affects the way in which business models ensure their sustainability. Therefore, this presents an additional challenge for data intermediaries for more inclusive data governance, as their business models not only have to be sustainable over time, but also need to be compatible with their main activities, purpose and governance structure.

In this section we analyse how data intermediaries can build sustainable business models in the light of the different components that make up a business model.

3.1 Defining a business model

We define a business model as ‘the principles and mechanisms according to which an organisation creates and delivers value to stakeholders while ensuring the conditions of its own long-term reproduction’ (Coriat, 2022).²² Note that the use of the term ‘value’ is deliberately wide in this definition. It refers to anything that a physical or legal

22. The translation is ours.

person considers valuable because it satisfies a need, be it economic (e.g., a monetary return, an increase in productivity, etc.) or not (e.g., enhanced privacy, the satisfactory feeling of contributing to a cause that matters, etc.). For example, a data union can create value by giving its members the capacity to negotiate the terms on which their personal data is governed by a platform. Note that, as we will explain in further detail below, a service might be perceived to be valuable by individuals or organisations even if they are not willing to pay (much) for it. Consequently, as it will be shown below, while creating value to stakeholders (creating ‘a value proposition’, cf. [Section 3.2.1](#)) is a necessary condition for a business model to exist, it is not a sufficient one. In order to ensure the conditions of its own long-term reproduction, the organization needs, among other things, to generate revenue streams (cf. [Section 3.2.4](#)) that transform its value proposition into income. Moreover, despite the use of the word ‘business’, the concept of ‘business model’ can be applied to any type of organisation, be it for-profit or not. In the case of the latter, even if the goal of the organisation is not making profit, its activities require engaging resources to carry on key activities in a sustained way over time, and probably engaging in monetary expenses vis-à-vis third parties. Hence, the organisation needs to devise a (not-for-profit) business model in order to ensure its economic sustainability.

A business model comprises several components. For the purposes of this report, we will focus on the four that are the most relevant to understanding the economic sustainability of data intermediaries:

- **Value proposition and key activities:** the customer’s need(s) met by the organisation and what the organisation does in order to satisfy such needs.

- **Key resources:** the physical, intellectual, human and financial resources used to carry on the key activities.
- **Costs:** the monetary costs which the organisation incurs to carry on the key activities.
- **Revenue streams:** the monetary influx that the organisation receives.

Let us exemplify with the case of a data intermediary that does nothing but to intermediate between data users on the one side, and companies and individuals on the other side. The **key value proposition** offered to both parties is intermediation: the intermediary facilitates both parties to find each other and ease the transmission and management of data access and use for them. Its key activities are technical and business oriented: generating a software solution to manage and send data, finding interested companies and individuals, engaging them, etc. To do so, the data intermediary has some **key resources** such as its specialised employees, headquarters, its own software, etc. These resources imply incurring costs such as wages, rents, licences, equipment etc. In exchange for the service provided, the intermediary generates **revenues**. Revenues can come from different sources.²³ The intermediary can charge one or both parties²⁴ for the intermediation service provided under different pricing schemes (e.g., fixed fee, commission on the sale of a dataset, etc.). Furthermore, the organisation could be funded by donations from public and/or private organisa-



23. This illustrative example only considers the case of a data intermediary that only does intermediation. If we consider intermediaries outside the meaning of the DGA, so we also account for services that do more than pure intermediation (e.g., providing a data driven service such as data analysis), they can charge any of the two parties or even a third party for that service.

24. Note that a party refers to a type of agent (e.g., a company), and not to a specific agent (e.g. company A). For example, it is common for marketplaces to charge a commission to sellers but not to buyers. In this case, we would say that the party charged are all the sellers present in the marketplace.

tions (as, for example, Wikipedia and Signal) that, even if they are not (always) data users themselves, benefit from the increased sharing of data that the intermediary brings about.

The business model will be sustainable if the data intermediary manages to design it in such a way that certain conditions are met. First, it needs to design an attractive value proposition for at least one party. Second, it needs to devise revenue streams to monetise that value proposition. Third, the revenue streams have to be at least equal to the costs that the data intermediary incurs in order to deliver its value proposition. Fourth, the key resources needed to deliver the value proposition need to be harnessed over time.

In the following subsections, we describe each of the key elements of a business model as described above with a focus on how they materialise in the particular case of data intermediaries.

3.2 Conditions of sustainability of data intermediaries

In this subsection we describe the most important conditions that have to be met for data intermediaries to find a sustainable business model. Although we tackle each of the four factors, we delve deeper into two conditions: building an attractive value proposition and harnessing key resources in the long run. We argue that these are the two major factors that explain the difficulties data intermediaries currently face when scaling-up.

3.2.1 Value proposition: building an attractive value proposition on key activities

From a business model perspective, we can see that data intermediaries are defined by

one of their key activities: they ‘connect individuals and companies on one side with data users on the other’.²⁵ In the next lines we detail what are the types of value proposition a data intermediary might have. These are not mutually exclusive. On the contrary, a data intermediary can combine multiple value propositions, which should raise its probability of finding a business model that ensures its economic sustainability.

Given that data intermediaries necessarily intermediate between at least two parties (i.e., data subjects and data users), they need to create a value proposition for both of them. In other words, they need to provide them with incentives to engage in data sharing through a data intermediary:

- Data subjects need to be given an incentive (e.g., a payment, a data-driven service, personal satisfaction on how their data is being re-used, etc.) that compensates them for their monetary and/or non-monetary (e.g. time and effort) costs of making their data available to re-users through an intermediary.
- Data users, in turn, also need to be given an incentive to use data intermediation services rather than not using data or obtaining it by other means. These incentives can sometimes be of the same nature (e.g. both parties can benefit from a data-driven service when data is shared) or not (e.g. the data subject enjoys the moral satisfaction of being able to share its data for a good cause, while the data user solves a practical problem with the data).

25. <https://digital-strategy.ec.europa.eu/en/policies/data-governance-act-explained>

3.2.1.1 Intermediation: reducing transaction costs

The mere matchmaking between individuals and companies with data on one side, and data users on the other side, can only constitute a value proposition if the transaction costs are sufficiently high. In other words, if it was easy and costless for a data holder and a data user to meet, agree on the terms of data sharing, and carry on with the exchange (for example, because they know and trust each other and they both use the same standardized technical protocols and data sharing licence), they would not need a third party for data sharing to happen. Data intermediaries can reduce **different types of transaction costs**.

- First, they can **reduce search costs**. By bringing data holders and data users together, and by adding certain features (e.g., tags, search bars, description of the datasets, sample datasets, etc.), data intermediaries can facilitate the matchmaking process, and hence reducing the costs, time and efforts, that would have been used for finding the data.
- In the same vein, they can **reduce information costs**. Each data holder/user needs to provide information to facilitate the exchange to the data intermediary once. Then, the latter acts as a one-stop-shop where data users/holders can find all the potential parties with whom they could share data and the relevant information to do so. This saves them time and resources that they would have otherwise spent in finding possible mutual gains in exchange with other data holders/users.
- Another way in which a data intermediary can reduce transaction costs is by **reducing bargaining-related costs**. Features such as standard

clauses for data licencing and standardised pricing structures reduce the possible options that data holders and data users face before reaching an agreement. They also increase legal certainty, as the parties reduce the time spent devising and protecting against possible legal issues generated by data sharing when seeking an agreement.

- Finally, a data intermediary could **reduce monitoring and enforcement costs** if it has a way to undersee how data is used and, in that way, detect potential breaches of the agreement. Importantly, by increasing legal certainty and providing monitoring and enforcement of the data sharing contracts, data intermediaries can also increase the level of trust between data users and data holders, which favours data sharing.

For example, suppose that company A is willing to make its data available to any potential user, and that company B, in turn is willing to use it and provide an economic compensation for it. Suppose that there is a price that both companies are willing to accept. In this case, the sale would not happen if, in the absence of an intermediary, search or transaction costs were too high. An example of search costs could be that company B does not know that company A has the data it needs. It would be simply too costly for company B to contact every company that might have the data it is looking for (provided that company B knows which companies these are and can obtain their contact information) until it finds company A. The same is true for company A, who wants to monetise its data. If, on the contrary, companies A and B know there is an intermediary that knows all the compa-

nies interested in ‘selling’²⁶ or ‘buying’ data, recourse to it would be much simpler. In this case, the data intermediary’s value proposition would be to reduce search costs. Hence, at least one of the two companies would be willing to pay the data intermediary.

The intermediary could also have a value proposition based on reducing bargaining-related costs. If companies A and B knew each other (e.g., because A is a supplier of B) and had an idea of the datasets each has and needs, they could still refrain from the data sale because transaction costs could be too high. For example, they could lack standard contracts fit for the type of data to be sold and the purpose of their re-use. In that case, an intermediary that proposes standardised ‘selling’ procedures and conditions could reduce both companies’ transaction costs, which would constitute a value proposition for which at least one of the two companies is willing to pay.

Note that reducing transaction costs constitutes a value proposition for both the data subject and the data user. Whatever the benefit (be it economic or not) the data subject obtains from sharing its data with a data user, it will only do so ‘through a data intermediary’ if the latter facilitates this process sufficiently that it prefers to use its services rather than bypassing the intermediary or desisting from sharing the data altogether. Conversely, whatever the data user’s reason for wanting to use the data, it will only choose to do it through a data intermediary if the latter reduces transaction costs sufficiently. Otherwise, it would simply desist from using the data, or obtain it in some other way.

26. Hereafter, we use the term ‘sell’ in inverted commas to signal that we are not using it in a legal sense. In the case of personal data, the GDPR excludes the possibility of considering it as property. Hence, individuals cannot give away all their rights over their personal data. In the case of non-personal data, what is ‘sold’ is, in most cases, a licence to use data, not the data itself.

3.2.1.2 Aggregating data

When data intermediaries connect data holders with data users, the latter are usually interested in an aggregated dataset made of data from several data holders. This is because datasets with more observations are subject to economies of scale in data aggregation (Floridi, 2014; Batini & Scannapieco, 2006; Olson, 2003; Wang, 1998), and datasets with more variables about the same observations are subject to economies of scope in data aggregation (Carballa-Smichowski et al., 2022).

For example, suppose a pharmaceutical company wants to access data about genetic information in order to develop a new drug. Having data about only a few individuals would not be enough to make statistically-robust inferences that could lead to a successful new drug. Moreover, even if the company had information about millions of individuals, if that information was limited to a few characteristics (e.g., sex and age), the number of variables about the observations (i.e., the individuals) would be too low to develop the drug. The pharmaceutical company might also need data about the physical activity, the eating patterns and some key symptoms of these people in order to successfully develop a drug to treat a genetic condition. If a data intermediary was able to obtain the consent from all those individuals and interphase with the different data holders that control access to the different variables of interest (e.g., a medical wearable company, a hospital, etc.), it could constitute an aggregated dataset of interest to the pharmaceutical company. In that case, the data intermediary’s value proposition would consist in aggregating data.

3.2.1.3 Increasing data quality

An intermediary could also provide an additional service consisting in increasing the (aggregated) dataset's quality. Data quality can be defined in multiple ways (Floridi, 2014; Batini & Scannapieco, 2006; Olson, 2003; Wang, 1998). For the purposes of this report, let us simply state that data quality comprises the following properties: accuracy, objectivity, accessibility, security, relevancy, timeliness, interpretability and understandability (Floridi, 2014). Examples of other dimensions of data quality are interoperability (how easy it is to use the data with other data or a software), discoverability (how easy it is to find the data), and its ethical quality (e.g., in the case of personal data, knowing that individuals are aware and have agreed to share it). Regarding interoperability, it is worth noting that the Data Governance Act encourages data intermediaries to transform data into a different format to enhance interoperability.²⁷

The relevance of each of these properties depends on the intended use of the data. Once the data intermediary obtains the data from the data holders, it can increase their quality in multiple ways, for example, by homogenising categories, cleaning the dataset of mistakes, etc. This adds value to the dataset around which sharing the data holder and the data could establish a commercial relationship, especially if only few enterprises are able to clean the data. Therefore, it makes data sharing (the commercial relationship) more interesting to the data user. This increases its willingness to pay for a data intermediation service. Increasing the quality of data can therefore

27. In particular, Article 12(d) of the DGA states that digital intermediaries 'shall convert the data into specific formats only to enhance interoperability within and across sectors or if requested by the data user or where mandated by Union law or to ensure harmonisation with international or European data standards'.

contribute a value proposition for which the data user is willing to pay.

3.2.1.4 Providing a data-driven service

Another way that **(non-DGA-compliant)** data intermediaries can provide a value proposition consists in providing data-driven services to the data subject/holder, the data user or both.

For example, many Personal Information Management Systems (PIMS) not only provide an intermediation service between individuals and third parties, but also provide individuals with technical tools to ensure their privacy is respected. To some individuals, a service that provides privacy protection constitutes a value proposition they would be willing to pay for. Other examples of data-driven services could be data analytics. A data intermediary could process and analyse the data to be able to supply them to clients in a more summarised form, rather than transmitting a raw dataset.

However, to the extent that the Data Governance Act's definition of data intermediaries as 'neutral' third parties implies that they cannot engage in the provision of additional services that overlap with that of intermediation, the capacity of data intermediaries to provide a data-driven service (and hence to provide an attractive value proposition) is constrained (see [Box 3](#) below). If an entity wishes to also offer data-driven services and aims to make economic benefits from that activity, two separate legal entities must be constituted under common ownership: a pure data intermediation legal entity (the data intermediary) and (a) data-driven service provider(s).

Box 3. The concept of neutrality in the Data Governance Act

The Data Governance Act only concerns ‘neutral’ data intermediation services providers. The term ‘neutral’ refers to two cumulative criteria data intermediaries have to meet in order to have to register according to the DGA:

- **Structural separation.** A legal entity considered to be a data intermediary can only provide data intermediation services. Hence, if it provides data-driven services beyond intermediation such as data analytics, it should do so through a separate legal entity.
- **Non-exclusivity.** The data intermediation service provided by a data intermediary should be open to any third party that respects the terms and conditions of the intermediary and the legal framework.

If these criteria do not apply, the entity will not have to register and thus will not be recognised as a neutral data intermediation services provider in the EU and able to register under the DGA.

3.2.1.5 Data governance as a value proposition

In some cases, the mere capacity a data intermediary gives data subjects to govern their data constitutes a value proposition. Some individuals care sufficiently about privacy or agency over their data, but lack the capacity or time to protect or exert them. If a data intermediary gives them the technical capacity to easily protect the privacy of their data and choose who to share it with and under what conditions, this constitutes a value proposition to these data subjects.

3.2.2 Key resources: harnessing key resources in the long run

In order to carry on with its key activities, an organisation needs resources. These can be physical (e.g., land), intellectual (e.g., a patent, a brand, etc.), human (e.g., a team’s expertise on a matter) and financial (e.g., venture capital or a bank willing to finance the production process) resources used to carry on its key activities. This is the case both for for-profit and not-for-profit organisations. For example, a data union without a commercial goal needs a legal team and individuals (typically volunteers) to carry on

advocacy, negotiation and technical tasks such as auditing automated data governance practices. In this case, these individuals constitute the key resources of the data intermediary.

For data intermediaries such as data cooperatives or data unions, in which data subjects are also key resources, harnessing the latter in the long run can also be very challenging (see [Section 3.1](#)).

3.2.3 Costs: controlling costs

Carrying on key activities implies costs. Even in cases in which a data intermediary is a not-for-profit organisation relying on voluntary labour, it still has to buy inputs, capital goods and services from third parties to operate. Although minimizing costs is important in any business model, a business model can be cost- or value-driven. In the first case, minimizing costs in order to be able to offer a low-price service is the driver of the organisation’s competitive advantage. Low-cost airlines are an example of this case. In value-driven business models, the organisation focuses on value creation and is less concerned about the cost implications of offering a unique service.

Data intermediaries, at least in their current starting stage, fall into this category. Currently, there is no established market on which data intermediaries compete fiercely. Moreover, users are not familiar with ‘standard’ data intermediation services. Hence, data intermediaries’ cost structures are value-driven. Data intermediaries’ focus is on first developing an attractive service that stands out (e.g., a marketplace specialised in transportation data, a data trust to manage private-public data from different stakeholders in a city, etc.), and then on minimizing the costs, rather than on prioritising low costs.

When having to minimise costs, the existence of economies of scale is a particularly relevant feature of data intermediaries to be considered. Economies of scale stem from a high proportion of fixed costs over total costs. Like many digital services, data intermediation services require the investment of considerable resources in capital goods such as servers and software development. Importantly, these expenses are needed irrespectively of the number of users. With digital goods, providing the service to an additional user has a small additional cost.²⁸ Therefore, in order to succeed in minimizing costs, it is key for data intermediaries to obtain a critical mass of users that allows them to recoup their high fixed costs, thus harnessing economies of scale.

3.2.4 Revenue streams: designing revenue streams

Once a value proposition materialises, data intermediaries need to determine which stakeholders could pay for it, how much and how. For the business model to be sustainable, revenue streams have to be at least equal to costs.

28. The fact that digital goods have a small marginal cost of production has been widely established in the economics literature. See for example Belleflamme (2016), Pénin, J. (2015) or Wunsch-Vincent, S. (2013)

Although users of the services, at the core of the value proposition, are typically the stakeholders that pay, it is possible for other parties to pay, too. If data exchanged through an intermediation service or produced through a data-driven service, like analytics, has value for a third party not directly engaged in the intermediation (e.g., a company interested in re-using the data to develop a product, a government that uses the data to implement or design a regulation), the information it provides can constitute a revenue stream. For example, a non-DGA-compliant data intermediary could aggregate data of its users, such as data subjects, and sell access to data analytics on different aspects such as mobility (e.g., to local governments trying to redesign the routing of a bus network), health (e.g., to researchers) or energy consumption (e.g., to companies providing customised energy contracts to households so as to reduce their energy bills). If this revenue stream is sufficiently important, individual users could be not charged to use the service. This creates incentives for them to use the service, and therefore to provide their personal data that ultimately create a revenue stream. The most widespread form of this practice is through targeted advertising, however, both for data intermediaries for inclusive data governance and for DGA compliant DISPs, a revenue stream through targeted advertising is not desirable and should not be pursued.

Once the stakeholders paying for the value proposition have been determined, the pricing mechanisms can be designed. These include, but are not limited to, subscriptions with the possibility of including freemium subscriptions, where users pay only if they want additional features), pay-per-use, progressive pricing based on the type of services or amount of data handled, intermediation fees (especially for marketplaces), donations, etc. The pricing mechanism is ultimately a data-intermediary-specific choice for which there is no best practice to implement.

4 The landscape of data intermediaries

Although initial discussions about data intermediaries date back to the early stages of the Internet (in particular the conversation around Personal Information Management Systems), data intermediaries are still a rather new concept and the terminology is in flux. Conscious of the different definitions adopted by academics, practitioners and policy makers, in this report we describe six types of data intermediaries that, to different extents, promote a more inclusive data governance, with the goal of providing greater clarity about their main features, governance mechanisms and business models.

In drafting this section, we would like to acknowledge previous attempts at the classification of (emerging) data intermediaries (e.g., EPRS, 2022; UK AI Council & Ada Lovelace Institute, 2021). Building on previous work, the goal of the report is to identify a **common understanding of different types** of data intermediaries, which can foster more inclusive approaches to data governance, and to contextualise these within the **regulatory (DGA) and economic context**. Not all types of data intermediaries mentioned in this report equally foster inclusive approaches to data governance. They do so in different ways and degrees, ranging from the promotion of greater control over data through more horizontal relations between actors, and value production from data reuse for the benefit of more actors and society.

The following six types of data intermediaries are described in this section: Personal

Information Management Systems (PIMS); data cooperatives; data trusts; data unions; data marketplaces; and data sharing pools. They differ according to various parameters such as objectives, governance structures, value propositions and the types of organisations involved. Yet they are not completely distinct from one another and might overlap in certain aspects (see [Table 2](#)). At the current time the landscape is still dynamic and governance practices have not yet crystallised - definitions should therefore still be flexible and, to a certain extent, open. The labels and descriptions provided below may be subject to further changes, depending on the developments that might occur in the field.

The value of this review lies in its attempt to provide a concise and multidisciplinary distillation of the key features of a wide range of data intermediaries that are commonly mentioned in this field and which can potentially foster more inclusive data governance in Europe and beyond. Furthermore, the section provides background knowledge about the overall landscape of data intermediaries for the stakeholders concerned with DISPs, i.e., data intermediaries complying with the requirements of the DGA that will have to register starting from September 2023. To better understand how the six types of intermediaries described in this section could be associated with DISPs covered by the DGA, see below ([Box 4](#)).

The choice of these six types has been made so as to consider intermediaries that are frequently mentioned in the literature

(while also being relevant in the context of the DGA), and that can contribute to promoting a fairer data ecosystem, by strengthening the power of data holders and data subjects and by fostering data sharing for both economic growth and societal wellbeing. This explains why intermediaries like data brokers are not included: data brokers are established with the main goal of aggregating and analysing data in order to monetise it by a single private company, and this has nothing to do with promoting inclu-

siveness in the realm of data governance.²⁹ Moreover, we did not include in this report data collaboratives or data commons, as they are different concepts, encompassing a wide range of data relations (the former), or assigning different kinds of rights over data (the latter), and neither specifically fits the category of data intermediaries.

29. Regarding data brokers: *One important distinguishing factor between data brokers and data market providers is that data brokers are actively engaged in the collection of additional data and their aggregation, while data market providers are passive intermediaries through which data controllers, including brokers, can offer their data sets.* (OECD, 2019)

Box 4. Correspondence between types of data intermediaries in Chapter III of the DGA and types described in this report

Data Governance Act - Article 10	In this report
a Intermediation services between data holders and potential data users; this concerns entities that allow bilateral or multilateral exchanges of data, the creation of platforms or databases enabling the exchange or joint use of data, as well as the establishment of other specific infrastructure for the interconnection of data holders with data users.	<ul style="list-style-type: none"> ● Data sharing pools (4.6) ● Data marketplaces (4.5)
b Intermediation services between data subjects or individuals that seek to make their personal or non-personal data available, and potential data users.	<ul style="list-style-type: none"> ● Personal Information Management Systems (PIMS) (4.1)
c Data cooperatives, which are organisational structures constituted by data subjects, one-person undertakings or SMEs. These entities will help members of the cooperative to exercise their rights over their data.	<ul style="list-style-type: none"> ● Data cooperatives (4.2) ● Data trusts (4.3) ● Data unions (4.4)

The Recitals of the DGA also mention specific types of data intermediaries, such as data marketplaces; orchestrators of data sharing ecosystems (data spaces) and data pools (Recital 28), which could be examples of type a, as in the table above; and personal data spaces (Recital 30), which could be an example of the type b above.

Each of the sections that is dedicated to a particular type of data intermediary is structured as follows:

- **Introduction:** a brief description of the specific type of data intermediary, its main objectives and the key stakeholders involved in its use.
- **How they work:** the main features and governance mechanisms that characterise the particular type of intermediary.
- **Examples:** selected pioneering, well-known or established examples of the type of data intermediary. The examples do not claim to be representative. They have been selected by the authors and verified by experts who read an earlier draft of this report to provide an overview of possible use cases. Although validated reviews or institutional databases of data intermediaries do not exist yet, we were able to draw from significant work carried out by MyData Global (Langford et al, 2022) and by other organisations (Mozilla Foundation,³⁰ The GovLab,³¹ Nesta,³² the Open Data Institute (ODI)³³ and the Ada Lovelace Institute³⁴), which have organised into categories existing operators, new data governance models and data partnerships.

- **Business model considerations:** The key issue(s) specific to the particular type of data intermediary, which need to be addressed in order to be able to design an economically-sustainable business model.

30. Database of initiatives. Alternative Data Governance in Practice <https://foundation.mozilla.org/en/data-futures-lab/data-for-empowerment/who-is-innovating-database-of-initiatives/>

31. Data collaboratives explorer <https://datacollaboratives.org/explorer.html>

32. The New Ecosystem of Trust <https://www.nesta.org.uk/blog/new-ecosystem-trust/>

33. The Data Institutions Register <https://theodi.org/article/the-data-institutions-register/>

34. Exploring legal mechanism for data stewardship and Exploring principles for data stewardship <https://www.adalovelaceinstitute.org/project/exploring-principles-for-data-stewardship/>

Table 2. Overview of the six types of data intermediaries covered in this report

Type	Main goals	Key stakeholders	Value proposition / business models considerations?	Type of approach	Distinguishing factors*
PIMS	Provide tools to individuals to take control over their personal data	Individual data subjects	Finding alternative revenue streams to user payments (e.g. data-driven services)	Individual	Technologies for individuals to manage the use of their personal data
Data cooperatives	<ul style="list-style-type: none"> Establish a bottom-up democratic governance structure Produce benefits for the members of a community 	<ul style="list-style-type: none"> Communities Members of cooperatives Individual data subjects Individual data holders 	Providing incentives for individuals to devote time to the different tasks they carry out	Collective	<ul style="list-style-type: none"> For specific purposes or groups, not for universal scaling. Based on democratic principles through the production and management of common pools of data
Data trusts	Establish responsible data management through independent decision processes in the interest of data subjects/holders	Individual data subjects	Focusing on increasing uptake in order to generate sustainable revenue streams	Individual	For delegating the management of rights over personal data to a trusted entity
Data unions	Establish collective bargaining on rights to personal data generated through platforms	Individual data subjects belonging to the same social group category (digital workers, platform users)	Providing incentives for individuals to devote time to the different tasks they carry out	Collective	For increasing bargaining power of platform users and digital workers, based on democratic principles and workers' dignity
Data marketplaces	Match data supply and demand	Business data holders	Reducing search costs and facilitating data exchange initiation between supply and demand.	Business oriented	Match supply and demand among data holder companies or entities
Data sharing pools	Leverage data synergies among stakeholders with complementary datasets	<ul style="list-style-type: none"> Business data holders Organisations 	Improving data products and services, creating new business opportunities, and tapping unexploited data value.	Business oriented	Create common pools of data among data holders often creating partnerships among multiple partners

Source: JRC own elaboration with contribution of Viivi Lähteenoja

4.1 Personal Information Management Systems (PIMS)

4.1.1 Introduction

Personal Information Management Systems (PIMS) (Janssen & Singh, 2022b) are a kind of intermediary for the management of personal data that appeared in scientific literature nearly 30 years ago (Barreau, 1995). They have been named in different, though similar, ways over the years to distinguish some nuances; these labels include: personal information management services (Ctrl-Shift, 2014); personal data spaces (Lehtiniemi, 2017; Lähteenoja, 2023); personal data management systems (PDMS) (Abiteboul et al., 2015); personal data stores (PDS) (deMontojoye et al., 2014); personal data servers (Allard et al. 2010); personal data wallets (Tengberg, 2013); and personal data clouds (Anciaux et al., 2019).

PIMS refer to a series of technologies developed to offer data subjects a means to leverage control of the processing of their data. Their aim is to provide an alternative approach to data processing by increasing the possibility for individuals to control and steer the processing of their data. They can be categorised under the umbrella of the so-called privacy-enhancing technologies (PETs), defined by Janssen et al. (2020) as ‘technical tools and measures that can assist in addressing privacy concerns’.

PIMS have traditionally been understood as a new type of intermediary developed to provide technical means to leverage data portability for individuals, a right protected under Art. 20 of the GDPR.³⁵ The objectives of PIMSs are to capture, aggregate, store, and manage personal data originating from

different sources. They are tools for managing personal data (often including the possibility to conduct analytics and computation) and for monitoring or controlling the transfer of data to third parties (Janssen et al., 2020).

This type of intermediary is targeted at individuals (data subjects), unlike other data intermediary initiatives whose main objective is to foster business-to-business (B2B) data sharing or public interest purposes. In detail, as Urquhart, Sailaja and Mcauley (2017) highlight, ‘PIMS/PDS [Personal Data Spaces] are emerging as promising platforms to give data subjects (consumers) more control over their personal data and thus to restore user agency, including in the context of the Internet of Things’. Moreover, PIMS/PDS can enhance trust in data reuse, and function as an ‘information trust bank’, as they are able to assess and confirm the reliability and trustworthiness of data users (OECD, 2019, p. 39).

Before going deeper into the details of the model, it is worth mentioning that some specific initiatives that fit the category of PIMS are being developed. Their aim is to incentivise the flourishing of PIMS ecosystems, and they include the following:

- MyData Global³⁶: a movement set up ‘to empower individuals by improving their right to self-determination regarding their personal data’. A way of achieving this is through the so-called MyData Operators, personal data companies that have been recognised for their efforts to provide human-centric solutions to individuals for the management of their personal data. In 2022, 33 initiatives had been acknowledged.³⁷

36. <https://www.mydata.org/>

37. The complete list of awarded initiatives is available here: <https://apply.mydata.org/gallery/mPLRwJzJ>

35. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679>

- ANewGovernance³⁸: an international association of public bodies, associations, academics, start-ups, and corporations, that helps build a governance framework for sectoral Data Spaces with a focus on their personal data dimension. The association defines as ‘a global human-centric personal data network in which all organisations take part and where data can easily flow from one organisation to the other under the person’s complete control and transparency’.

4.1.2 How they work

In general, the technology behind a PIMS consists in ‘equipping an individual (a PIMS user) with a device dedicated to managing their data’ (Janssen et al., 2020). The device is intended as a technical environment where the user’s personal data are located, providing mechanisms for the user to control both the data that flow in and out of the device, as well as the computation that occurs with them, generally through applications.

When data users need to perform computations with a data subject’s data, they require permission to carry out the activity, and can only do this with respect to a specific set of data, through an agreement with the data subject. Computation with data may be performed inside the PIMS, without any data flowing from the user’s premises to the developer’s premises. In some cases there could be computation on the user’s device with only the results transmitted to the developer, while other cases could involve the physical movement of data from the user to the developer. Whatever the situation, in a PIMS the user will always have a complete view of what happens to his/her data, with the possibility of restricting some (or all) the consents given at any time (Janssen et al., 2020).

38. <https://www.anewgovernance.org/manifesto/>

Key features of PIMSs according to the OECD (2019) are the following:

- they provide user-friendly technical means to allow a user to gather and access their personal data, usually through a virtual space/cloud;
- they allow a user to specify which data may be shared and which not, and to which third party data user/platform;
- they vouch for the reliability and trustworthiness of the platforms that act as data users;
- they have a direct value to beneficiaries: the value proposition is targeted for the data subject (not for the common good, nor for businesses).

4.1.3 Examples

- **CozyCloud**³⁹ through its CozyCollect application provides users with a catalogue of data holders from which to retrieve their personal data, including financial data (e.g., Paypal or bank transactions); administrative data (e.g., electricity data or consumption history, insurance contracts); music data (e.g., Spotify and YouTube); fitness data (e.g., Fitbit, Garmin, Google Fit); social media data (e.g., Facebook, Instagram, Twitter); and medical records. It offers to its users a decentralised storage space in which they can manage, retrieve and organise their personal data.
- **JoinData**⁴⁰ is a Dutch PIMS dedicated to farmers. It allows them — through the payment of an annual fee — to manage all the permissions about their company data in one single platform, as well as to share those data with different parties (government, suppliers, customers, etc.). In

39. <https://cozy.io/en/>

40. <https://join-data.nl/en/>

this case data subjects are not individuals managing their personal data, but farmers that want to manage their companies' data (data holders).

- **Mydex**⁴¹ is a data infrastructure for businesses and developers, which allows individuals to use their own personal data stores to collect, use and share personal data (Mydex, 2021). Mydex is organised as a Community Interest Company (CIC) in Scotland and its platform services are available to app developers (upon payment of a fee) on the UK Government Digital Marketplace, the 'G-Cloud'.
- **Bitsaboutme**⁴² is a tool that allows data subjects to regain control over their personal data with respect to their purchasing habits through loyalty cards. BitsaboutMe developed a secure platform for the fair exchange of personal data between consumers and organisations. The app allows consumers to collect and share anonymised information about themselves, and to receive in return reports about their purchasing behaviour and lifestyle. They can also share their data, earning real money or reinvesting earnings into offsetting their carbon footprint. From the organisations' side, the app allows them to benefit from high-quality, anonymous data from multiple sources that comply with privacy laws.
- **Citizen-me**⁴³ self-defines as 'the only Zero Data app in the world', meaning that users remain in full control of their data, which is stored in their device and not in the app. Users can decide to share data (defined as Me-Data) with consent; after having anonymised the data and removed any

personal identification, Citizen-me provides users with a fair reward for their data. This reward is composed of free personality tests aimed at improving health and life choices, as well as payments through PayPal after completion of a data exchange. The app also allows data to be donated to climate change research, charities, or medical research.

4.1.4 Business model considerations

Regarding business model design, one of the key challenges of Personal Information Management Systems (PIMS) that are targeted at individuals is to generate a revenue stream. The value proposition of Personal Information Management Systems that target individuals consists at least in offering them a one-stop shop to gather and manage their personal data. Although this certainly constitutes a value proposition, some individuals may not be willing to pay much money for it. Data users, in turn, have low willingness to pay for data about single individuals, as opposed to an aggregated and curated dataset about many individuals. Hence, as the aforementioned Bitsaboutme and Citizen-me examples illustrate, some Personal Information Management Systems are starting to integrate additional services to enhance their value proposition and, in that way, generate new revenue streams. In the case of Bitsaboutme, the value proposition is enhanced by offering users the possibility of receiving reports about their purchasing behaviour and lifestyle, sharing their data, earning money or reinvesting earnings into offsetting their carbon footprint. In the case of Citizen-me, the value proposition is enhanced by offering users free personality tests aimed at improving health and life choices as well as payments in exchange for sharing their data.

41. <https://mydex.org/>

42. <https://bitsaboutme/en/>

43. <https://www.citizenme.com/>

4.2 Data cooperatives

4.2.1 Introduction

Data cooperatives allow data subjects/holders to have greater control over their own information, exercising their rights with respect to data, making informed choices about their processing and steering their use according to their motivations, preferences and concerns. What distinguishes data cooperatives from other types of data intermediaries is that they foster a more democratic governance approach because the purposes and conditions according to which data are shared, processed and used are based on agreements between members (Blasimme et al., 2018).

Data cooperatives are a ‘collective type’ of intermediary: individuals manage their data for the benefit of the whole community, to make their voices heard and to support societal or civic causes that matter to them. Members can both manage already existing data and collect or generate new data; [see for instance initiatives based on citizen-generated data for policymaking (Meijer & Potjer, 2018; Ponti & Craglia, 2020)]. Besides pooling data in the cooperative, members negotiate the ways in which data are processed, being themselves responsible for stewarding the data (Ada Lovelace Institute, 2021).

Data cooperatives support bottom-up governance often with the aim of reducing asymmetries of power and monopolies for data collection and use. The data cooperative model is inspired by the cooperative movement established in the UK and France in the 19th century and still playing a pivotal role in many industrial sectors across Europe. Recently this long-standing model has also become increasingly popular in the digital economy, following the advent and development of platform cooperatives and data cooperatives (Scholtz, 2016; Borkin,

2019; Craglia et al., 2021). Like platform cooperatives, data cooperatives offer a positive alternative to Big Tech monopolies (Sutton, 2016) by attempting to rebalance the relationship between data subjects, data platforms and third-party data users, addressing the information and power asymmetries of the data economy.

The stakeholders leading a data cooperative can vary, and may be individuals, communities, civil society organisations, SMEs, research institutions and NGOs. Although many data cooperatives are ad hoc initiatives, centred around the new flows of data they enable, data cooperatives can also be built on existing cooperatives, for instance in the agricultural sector. Even if this is not often acknowledged when addressing data intermediaries, cooperatives are a distinct type of enterprise, different from private companies, NGOs and public bodies — and they can have either a for-profit, or a not-for-profit business model, depending on how they are implemented (Mannan et al., 2022). Leaning on existing (offline) cooperatives, helping them to add ‘a data layer’ could be a promising strategy to support the diffusion of this type of data intermediary. As proposed by recent work from the Aapti Institute, implementing data cooperatives on top of offline cooperatives can be more effective and socially relevant than building new data services, especially among unprivileged communities hit by digital divides and inequalities.⁴⁴

Data cooperatives have been flourishing in particular in the health sector, with the goal of enabling individuals to donate their personal health information for scientific research. The fact that many cooperatives concern health-related data is due to the tensions and disputes surrounding the governance of sensible personal information and the great potential it holds for



44. <https://aapti.medium.com/exploring-the-value-of-adding-a-data-layer-to-cooperatives-megha-farmer-cooperative-case-study-1c4fcfd08635>

research, from gene sequencing information to self-tracking data and administrative medical records (Beaulieu & Leonelli, 2022). In the section outlining examples

below (4.2.3), we report two data cooperatives from the health sector and three from other sectors.

Box 5. The Data Governance Act provisions on data cooperatives

In the DGA, data cooperatives are regulated as one type of ‘data intermediation service’. **Data cooperatives are defined under Art. 2(15) as data intermediation services offered by an organisational structure constituted by data subjects, one-person undertakings or SMEs who are members of that structure, having as its main objectives to support its members in the exercise of their rights with respect to certain data, including with regard to making informed choices before they consent to data processing, to exchange views on data processing purposes and conditions that would best represent the interests of its members in relation to their data, and to negotiate terms and conditions for data processing on behalf of its members before giving permission to the processing of non-personal data or before they consent to the processing of personal data.**

4.2.2 How they work

Data cooperatives offer tools for exerting direct control over (personal or non-personal) data, participating in the governance of the cooperative and pursuing collective benefits. In principle, cooperatives foster a ‘communal approach to data sharing’, intended as a ‘decentralised model’ in which **the whole community participates in decision-making**, such as when data is managed as commons (Ho & Chuang, 2019). The community of a data cooperative should be able to decide on several matters, such as the rules, norms and principles for data use; how data are collected; for what purposes the data are used; and who can access the data, etc. Data subjects can preserve greater control over their data compared to what happens in most other data governance regimes and might also receive an equitable share in the benefits produced from the use of the data (Borkin, 2019; Delacroix & Lawrence, 2019).

Data cooperatives often aim to create public value, for instance by addressing societal issues like environmental causes or medical

research. For such a reason, they can also be ‘commons-based’ and open (‘open cooperativism’), when data are shared with an open licence and made public (Carballa-Smichowski, 2019; Sandoval, 2019).

Key features of a data cooperative include the following:

- Data cooperatives are free associations and communities of individuals or data holders that steward data in the interests of their members, allowing them to have greater control over the data they generate.
- Data cooperatives are **democratically and collectively managed**. Members of a data cooperative take part in the governance of the data, either directly or indirectly, by developing governance policies or electing members of oversight committees. Participatory governance methods might be adopted to define the conditions for data reuse so that members can have a role in designing licences and terms and conditions for data access and use by third parties.

- Data cooperatives are based on the collaborative collection or retrieval of data owned by their members. They often draw on existing data, which is gathered as a new cooperatively governed data asset. Often, but not necessarily, they explicitly define data as commons managed by and for the community in terms of who owns the data and how it is governed. They are not based on promoting individual ownership and control, but, on the contrary, on the collective ownership of data (also intended as commoning).
- Data from cooperatives generate benefits for the members of the community that have shared interests. Members of data cooperatives have an equal share in the benefits and might be able to decide how revenues will be reinvested. Members seek to satisfy their collective interests (e.g., increased knowledge about rare diseases) that cannot be pursued individually.
- Data cooperatives steer data towards desirable data usage — they often pursue a public interest purpose, producing benefits for the wider society with members motivated by philanthropic purposes.
- Data cooperatives have an ambitious objective: their main goal goes beyond leveraging individual rights over data, in line with data protection law — they are established to set up a governance infrastructure to allow people to collectively decide about how data is used and what rules apply. They aim to enable members to act as collective choice makers and take meaningful decisions about their data so as to release their relational value.

4.2.3 Examples

- **Salus.Coop**⁴⁵ is a not-for-profit citizens-led data cooperative to promote health care innovation and research, founded in Barcelona in September 2017. Salus.Coop creates a collaborative governance model for the management of health data, that allows members to decide what research they want to support and to be the managers of their own data. This data can be of different kinds (e.g., location data, information from devices and apps, official national health reports data, surveys, etc.). It ‘legitimises citizens’ rights to control their own health records while facilitating data sharing to accelerate public research innovation in healthcare’ (UK AI Council & Ada Lovelace Institute, 2021, p. 55). The licence through which data is shared was determined through a participatory process in which it was decided that it had only should be used for not-for-profit research with an open-source approach.
- The **MIDATA** cooperative⁴⁶ was created in 2015 by ETH Zurich (Computer Science Department and Institute of Molecular Systems Biology) and the Bern University of Applied Sciences (Institute for Medical Informatics). It is an example of a not-for-profit data cooperative enabling access to data for public interest purposes, specifically to support medical research projects. (Blasimme et al., 2018). Members can share their specific medical causes and projects according to their interests and those of the community. Decision-making is by means of democratic voting mechanisms in the context of a general assembly of cooperative members.



45. www.saluscoop.org

46. www.midata.coop

- The **Grower's Information Service Cooperative (GISC)**⁴⁷ is an American cooperative allowing farmers, on payment of a monthly fee, to collect, store and manage their data. Farmers can profit from the cooperative's partnerships with experienced technology companies to gain business insights about their activity. In 2019, this initiative was awarded the 'Ag Data Transparent' seal that provides farmers with transparency, simplicity and trust in their agricultural technology contracts.
- **SAOS**⁴⁸ is a Scottish cooperative working on farming and food production (operating since 1905), that also provides data cooperative services. It has implemented ScotEID, a data

system that allows farmers to register, track and share data on their livestock. ScotEID is a cooperative database that offers real-time information on livestock across Scotland, which is vital to prevent disease outbreaks and to enhance the traceability of the food supply chain. Being managed as a cooperative, all the data in ScotEID are owned and controlled by the member farmers who generate this data.

- **POSMO coop**⁴⁹ is a data cooperative for mobility data, based in Switzerland. Posmo was launched in 2020 with the goal of building a data market that allows interested purchasers to access a pool of mobility data which is created and administered by the data subjects.

47. www.gisc.coop

48. saos.coop/what-we-do/data/

49. <https://posmo.coop/>

Box 6. The Data Governance Act provisions on data cooperatives and data altruism organisations

Depending on how they are implemented in practice, specific use cases of data cooperatives might fall either under the label of 'data altruism organisations recognised in the Union' (RDAOs), or neither of the two.

The definition of data cooperatives in Chapter III of the DGA (as one type of DISP) mainly highlights how they can support the realisation of individual objectives for the participating members. On the other hand, from the literature and engagement with the external experts during a workshop, we have found a greater emphasis on their facilitation of collective data management and on public interest goals.

Since data cooperatives are often established to support data sharing for public-interest outcomes, and not for the purpose of establishing commercial relationships, they might also fall under the label of data altruism organisations recognised by the Union. If income is earned by sharing data with third parties, it is often used to ensure the sustainability of the initiative, so they would still qualify as data altruism organisations according to the DGA. However, data cooperatives also imply that members remain in control of the data, according to modalities defined by members of the community — which might not fully comply with the European data altruism consent form that is to be established by the DGA (to be seen case by case).

4.2.4 Business model considerations

Those data intermediaries in which stakeholders are directly involved in the governance of data, and where personal data are more relevant — such as data cooperatives or data unions — may find a challenge in terms of harnessing a key resource: data subjects. In these types of data intermediaries, data subjects need to devote time and effort for the key activities to be carried out. They have to take part in regular meetings and discussions to define and enforce the collective governance of the datasets; they have to handle authorisations and conditions of use of their personal data; and, in some cases, they need to collect data on a regular basis themselves. Therefore, in these types of data intermediaries, data subjects are both users and key resources of the organisation.

In order to maintain these key resources in the long run, a data intermediary has to **provide incentives for individuals to devote time to the different tasks they carry out** and that are vital for the functioning of the data intermediary. These incentives can be of different types, as outlined below.

On the economic side, a possible incentive could be a payment by reusers of an individual's data in return for the monetisation of data. However, the effectiveness of this incentive depends on the sector in which the cooperative operates. For instance, the possibility of paying individuals for sharing their health data has been widely criticised — experts in the field have warned about the risk of widening inequalities and of reducing altruism by luring people to sell their privacy, and have suggested instead that health data should be considered as collective property (Prainsack and Forgó, 2022).

A different possible economic incentive would be receiving a data-driven service

from data users. This incentive exists if the only way for data subjects to receive the service, or for its quality to be improved, is if they share the data through an intermediary, rather than directly with the data user providing the service.

Members of a data cooperative can also have non-economic incentives. This is typically the case in the health sector, where individuals spend time and effort in collecting and governing health data to advance research about a rare disease or for improving their own health.

It should be noted that these incentives are usually specific to a narrow community of engaged individuals. When trying to scale up and reach the general public, a data cooperative might face the challenge that many individuals are not as engaged as the core community around which it was originally created. This poses a challenge, as the cooperative might lack key resources to operate once it starts to scale up. Moreover, even if the data cooperative has a sufficiently large user base of engaged individuals, the incentives that drive their engagement need to be sustained over time. If, for example, the non-economic reward that individuals obtain is a one-off, but their dedication is needed continuously, there is a risk of disengagement by data subjects.

Data cooperatives face another challenge: like most data intermediaries, they lack a clear revenue stream. However, in the case of those data cooperatives whose main stakeholders are individuals, this challenge becomes more acute. Because their main goal is usually to contribute to a societal cause through the gathering and sharing of data, they refrain from selling services to the main users of the data. Moreover, although individuals may be interested in the project, they are not likely to pay a sufficient amount (e.g., in the form of an annual contribution). To achieve financial sustainabili-

ty, data cooperatives might share data for a fee with third parties — under non-exclusive licence agreements — and use a portion of the value generated to sustain themselves (Blasimme et al., 2018; UK AI Council & Ada Lovelace Institute, 2021). However, if the project is niche and lacks continuous input by data subjects, the income earned would be very limited and not sufficient to guarantee sustainability.

4.3 Data trusts

4.3.1 Introduction

The prevalent definition of a data trust is based on a legalistic approach and consists in setting up specific legal mechanisms to ensure the responsible and independent stewardship of data. In a data trust a relationship is established between two types of entities: an intermediary (the trustee), which takes on the responsibility to steward the other entity's data/data rights for their benefit (the beneficiary). A data trust, which is based on trust law, allows data rights holders to delegate control of their data to a trustee (World Economic Forum, 2022). It represents a legal mechanism that permits the rights of data subjects/holders to be pooled and the terms of use for the data determined in the suppliers' interests (Sadowski et al., 2021). The notion of data trusts leverages existing legal governance structures, such as trustees' fiduciary duty, to provide the public with stronger protection against privacy violations and the unethical collection and use of their personal data (Element AI & NESTA, 2019, p. 28). Under certain definitions, data trusts can also be based on contractual or statutory legal obligations [(Open Data Institute, 2018; Reed et al., 2019; Ada Lovelace Institute, 2021; Global Partnership on Artificial Intelligence

(GPAI), Aapti Institute, & Open Data Institute, 2021) cited in Janssen and Singh, 2022a)].

In jurisdictions without trust law, or an equivalent, certain institutions set up a board of trustees to obtain the same goal. Overall, a data trust is expected to be independent and not to have conflicts of interests. It upholds *its duty of care by doing no harm to its clients and uphold its duty of loyalty by not having any conflicts of interest. Under a more expansive definition, a (data) fiduciary upholds its duty of care by protecting and enhancing the individual's digital experiences and upholds its duty of loyalty by actively promoting the individual's interest. The digital fiduciary can be an individual or an entity, a private or public (governmental) body and, if private, a for-profit or not-for-profit enterprise* (World Economic Forum, 2022, P. 14).

4.3.2 How they work

The implementation of data trusts is limited, and knowledge about concrete cases of the use of this type of data intermediary is scarce. The notion of a data trust is adopted to describe a legal mechanism and data governance model, but very rarely does it refer to actually implemented initiatives. Yet it is only if a variety of data trusts — each instantiating a particular way of balancing risks and responsibilities with opportunities for data reuse — were to be established, that individuals would be able to select and switch between trusts based on their needs and preferences (Delacroix and Lawrence, 2019). If an ecosystem of data trusts were established, the public would be able to choose a data governance regime that reflects their privacy preferences and supports their values.

Key features of data trusts include:

- Data trusts are often based on trustees' fiduciary duty, a specific legal mechanism according to which a trustee should do no harm and promote the benefits of its beneficiaries with undivided loyalty. In jurisdictions where trust law is not available, an option is to set up a board of trustees.
- The data trust must be fully independent and without conflict of interests in its role.
- Data trusts increase individuals' ability to exercise their rights, expanding their choices about data use, enhancing possibilities to save data reuse for public benefit and strengthening bargaining power to negotiate about data (for an analysis of the conditions under which those joining a data trust may mandate the exercise of their rights, see Giannopoulou et al., 2022).
- In a data trust, beneficiaries (data subjects/holders) mandate or delegate decision-making power to the trustee through trust law legal mechanisms or via a board of trustees. Compared to PIMS and especially to data cooperatives, they put less burden on the members. Decisions are taken by the trust, and no lengthy negotiations are needed.
- Experimentations exist around developing civic data trusts, which are independent and qualified data trusts implemented at the civic level, often to address citizens', companies' and utilities' lack of trust in giving their data to local governments.
- Specific mechanisms for deliberation and consultation with beneficiaries could be built into a trust (UK AI Council & Ada Lovelace Institute, 2021). Especially for civic data trusts,

implemented by a city's local administration, data trusts should be subject to public dialogue, scientific oversight, and democratic accountability to recognise the plurality of interests beyond data, for instance by providing representation on their board of trustees (Element AI & NESTA, 2019).

4.3.3 Examples

- Three data trusts pilots⁵⁰ have been launched in 2022 by the **Data Trusts Initiative** (www.datatrusters.uk). The first is a place-based data trust in the small fishing town of Brixham, UK, and is concerned with environmental stewardship for the benefit of the community. The second pilot provides an alternative to traditional consent routes for women taking part in an NHS longitudinal study of women who have given birth in Scotland (<https://datatrusters.uk/pilot-projects-3>). The third pilot focuses on one million persons who opted out of sharing their NHS primary care data in 2022 to examine the potential inherent in offering the option of sharing data via a data trust instead (<https://datatrusters.uk/pilot-projects-2>).
- **PLACE Trust** is a legal data trust that collects mapping data in collaboration with local governments across the globe and stores it in a trust in UK. It is based on a membership fee; it aggregates data and uses it for public interest purposes. In accordance to its website: *All data produced in partnership with PLACE belongs to the government of each country. PLACE receives from each government an irrevocable, perpetual, royalty free licence to a copy of all*



50. <https://datatrusters.uk/pilot-projects-1>

data and its use by PLACE members through the PLACE Trust. That Trust will issue licences for use of this data by our members. It preserves neutrality through the separation between the body that oversees the management of data and the one that controls economic operations. It has a 'board of trustees', registered in the UK which focuses on data use, and a 'boards of directors', which is based in the US and focuses on how to sustain the operations of the trust (<https://www.thisisplace.org>).

- A report by the UK AI Council and the Ada Lovelace Institute (2021) presents an interesting mock case study of a **possible data trust for educational technologies**. In such a context, a data trust pulls together the data rights from students and parents, whose personal data is collected by an educational platform provider. Then, the data trust negotiates the 'terms of service to the benefit and limits established by the school, parents and pupils' (UK AI Council and Ada Lovelace Institute, 2021, 47); it can maximise the ability to derive useful information from such data and leverage its members rights to data portability and access, while also minimising any risks of data sharing.

Among the challenges for establishing data trusts is the complexity of the skills required for data trustees, as they must ensure the privacy and security of the beneficiaries' (personal) information, as well as managing data in their interests and maximising the benefits they can derive from data.

To function well, and provide benefits for its beneficiaries, a data trust needs the participation of a large number of people. So, similarly to other intermediaries for personal data (data cooperatives and PIMS), uptake is a key challenge for success and scalability, but individuals might not be aware of the existence of data trusts, or they may be unable, or disinterested in joining a data trust.

Nonetheless, as listed in a report by the UK AI Council and the Ada Lovelace Institute (2021: 44), there are several possible funding models for data trusts, which range from privately and publicly funded, to *charging a fee or subscription from data trust beneficiaries* (the individuals or data subjects) *in return for streamlining and/or safeguarding their data interactions, charging a fee or subscription from those who use the data* (organisations), *charging individuals for related services, a combination of the above*. Furthermore, incentives for organisations in engaging with data trusts might include 'reputational benefits, legal compliance and future-proofing data governance practices' (ibid: 45).

4.3.4 Business model considerations

As already stated, it is challenging to identify implementations and use cases for data trusts. The notion of data trusts adopted in this section mainly refers to a proposed legal mechanism, or to prototypes in urban settings. These characteristics (legal mechanism, lack of use cases, prototype stage) make it challenging to identify key business model considerations.

4.4 Data unions

4.4.1 Introduction

Data unions are driven by democratic principles and their main objective revolves around protecting and bargaining on rights over personal data generated through platforms, both by users and workers. Data unions are mainly targeted at individuals

engaging in digital practices and producing data as a by-product such as users of Big Tech platforms or platform workers (e.g., food delivery riders). Zingales (2022) defined data unions as ‘a coordination by individuals to improve the conditions under which they make their clickstream data (which could be considered as the output of their labour) accessible to third parties’.

The value of this kind of data (that are ‘socially produced’) lies in their relational nature, so ‘there are mechanisms inspired by those of social welfare that would allow people to collectively bargain over the rights to their personal data that could be justifiable and virtuous’. (Carballa Smichowski, 2019, p. 224).

It can be argued that data unions overlap with data cooperatives or data trusts, depending on the governance scheme adopted when implemented. Data cooperatives are often established to empower people and create value from pooling data of their members, while data unions also empower people (data subjects), but mostly in respect of a specific data holder/platform that has collected their data. In practice, data unions seem to ‘only apply by definition to relational personal data massively collected by platforms such as Facebook, Yelp, Uber or Airbnb’. (Carballa Smichowski, 2019). Zingales (2022) offers a distinction between the two types of intermediaries, arguing that in a data union, ‘individuals get together to represent the interests of a community, and coordinate the use of their data in order to improve their socio-economic conditions’, while a data cooperative is ‘an autonomous association of persons uniting voluntarily to meet their common economic, social and cultural needs and aspirations through a jointly owned and democratically run enterprise’. Data unions also share some analogies with consumer associations, as their core mission is to protect

the rights of members and negotiate better conditions for them. Cooperatives, on the other hand, are directly involved in the production of goods and services in the interest of their members.

The main activity of data unions is to collectively negotiate the terms of the governance of data on behalf of a collective of data subjects; this feature reveals the analogy with (labour) unions that collectively negotiate working conditions on behalf of workers. Here the concept of ‘data dignity’ can be introduced as an evolution of that of labour dignity (Savona, 2020). This concept arises as we experience the lack of a labour market for data generators (Posner and Weyl, 2018). This is where data unions can play their role: a revisit to the traditional method of collective bargaining and worker representation adapted to the digital ecosystem (Savona, 2020).

4.4.2 How they work

Data unions or similar organisations position themselves between their members and platforms according to a simple mechanism, that is:

- Data subjects grant the intermediary (the data union in this case) the exclusive right to use the data.
- The data holders’ platforms offering services (e.g., Facebook, Airbnb and similar entities), on the other hand, must negotiate with the data intermediary in order to be able to collect, use and commercially exploit the data of data subjects (Radicalx-Change Foundation Ltd, 2020).

This represents a break in the direct relationship between individuals and platforms that underlies the ‘free service for free data’ model and in which individuals do not have enough bargaining power to defend their interests.

Two types of data unions may exist: sector specific or territorially based (Carballa Smichowski, 2019). Examples of the first type are social networks, crowdsourced review websites, ride-hailing apps, etc. — while for the second type the coverage of the data union could be a city, a region or an entire country. In terms of the activities that they carry out, data unions can bargain over measures with the aim of collectively protecting their members' privacy, as well as of sharing data with third parties (Carballa Smichowski, 2019, p. 224).

RadicalxChange Foundation Ltd (2020) proposes that data unions (or coalitions) should be governed according to democratic rules.

Some of the key requirements to which this kind of data intermediary conform according to Lanier & Weyl (2018) are:

- Fiduciary duties: data unions should serve the best interests of their members.
- Quality standards: they should commit to high quality standards for evaluating and tracking the quality of data and incentivise users to provide high quality data.
- Inalienable provenance: data unions should not allow users' data to be used beyond users' control and should grant access to third parties only for a defined use (in other words, permanent data sales should be prohibited).
- Benefit sharing: they should return a fair share of the value of the data to its producers (data subjects).

4.4.3 Examples

- **Swash**⁵¹ is a browser extension that claims to allow data subjects to 'get power' over their data while they re-

51. <https://swashapp.io/>

ceive a share of the profits they generate. It gives data subjects the possibility to aggregate their web surfing data and then 'sell' it to interested parties. As declared on its website, it 'enables individuals to practice their data rights while celebrating their role as essential value providers in the ecosystem'.

- **TheDataUnion**⁵² aims at creating a collective representation of Internet users, helping them with tools for claiming or deleting their data, as well as exerting political pressure on governments (to get stronger laws) and on tech giants, especially Google and Facebook, to get better terms and conditions.
- **Worker Info Exchange**⁵³ is a data union specifically targeted at platform workers (like Uber drivers or Deliveroo riders). It is a not-for-profit organisation that aims at helping digital workers in accessing and gaining insights from the data that are collected about them at work.
- **Unbanx**⁵⁴ is a reward app that allows data subjects to benefit from the value of their bank transaction data. This initiative builds on the fact that on their own, a person's data do not hold much value, but when combined with other people's data they can become an attractive product for buyers to extract insights. Moreover, it builds on the principle that banks are 'selling' their users' data, and therefore the profits from this activity should be credited to those who created the data.

52. <https://thedataunion.eu/>

53. <https://www.workerinfoexchange.org/>

54. <https://www.unbanx.me/>



- **DIMO**⁵⁵ is a data union that focuses on vehicle data with the aim of helping drivers to get more from the data they produce while driving and contributing to a future where driving apps work for vehicle owners, not for the companies that build them. Moreover, DIMO allows developers to use its open-source platform and unlock the value of real-world data from the tens of thousands of cars that contribute to the network.

4.4.4 Business model considerations

The economic reflections on data unions are similar to those on data cooperatives that can be found in [Section 4.3.4](#).

4.5 Data marketplaces

4.5.1 Introduction

Data marketplaces are platforms that enable the matching of the supply and demand of data or data products/services. These platforms act as ‘neutral intermediaries’ in data flows as (i) they do not actively intervene in data value chains but solely facilitate the matching of supply and demand (Spiekermann, 2019), and (ii) the data intermediation service is open to any third party that respects the terms and conditions of the intermediary and the legal framework (see [Box 3](#) above). Hence, data marketplaces fall under the scope of Chapter III of the DGA and are considered a prime example of data intermediaries in that legal text.

Data marketplaces typically establish standardised licensing models and rules regarding access to and use of data products within their scope. These characteris-

tics can lower entry barriers to engaging in data trading to a broader set of data providers, provided that the conditions for joining these environments allow for it, e.g., the costs to join are affordable.

As in any other multilateral marketplace environment, the marketplace owner should safeguard the needs and requirements of both supply and demand actors. Thus, the conditions for joining and participating in marketplaces should be designed to provide incentives for all participants. Failing to keep the incentives on any of the market sides may cause the collapse of even the most well-performing platforms, due to network dynamics. Examples of known strategic failures in these multi-sided environments include the failure to share the surplus, a focus on profit rather than on achieving a critical mass of providers, and the failure to optimise ‘openness’ to attract a larger number of actors (Van Alstyne et al., 2016). Integrating inclusive data practices, such as engaging in more participatory decision-making processes, could help achieve a successful uptake of data marketplaces.

Data marketplaces play a role in lowering the barriers to finding and accessing data assets for a broader set of actors. Specifically, the improvement of data findability and the standardisation of licensing models and access rules can enhance the conditions for data-driven innovation. Additionally, this type of intermediation can potentially add a layer of transparency to the bargaining conditions between data suppliers and data users.

According to the literature, no data marketplace has yet become a significant player in the data trading market (Koutroumpis et al., 2017). However, there are several ventures of these multilateral data trading



55. <https://dimo.zone/>

environments (e.g., xData, Kasabi or Pool⁵⁶) and there are also examples of active sector-specific data marketplaces, such as API-AGRO⁵⁷ in agriculture and Nxtport⁵⁸ in logistics (see Section 4.5.3).

4.5.2 How they work

Data marketplaces can take rather different governance structures. For instance, according to Spiekermann (2019) they differ in relation to:

- their value proposition: either transaction-based or data centric;
- their access mode: closed, hybrid or open;
- whether they are domain specific;
- their architecture: centralised, hybrid or decentralised;
- their pricing model: free, subscription, package, pay-per-use, progressive pricing;
- their revenue model: free, freemium, flat rate, fee, listing fee, transaction fee/commission, service fee, storage fee.

As with any other type of digital platform, the success of data marketplaces depends first on reaching a critical mass of both providers and users, and then on the replication of desired interactions — matching data suppliers with data consumers — to ultimately thrive through achieving economies of scale and scope.

Data marketplaces can also include platform orchestrators and third-party service providers, such as those providing services for facilitating the dissemination of data offers, and offering enhanced matching algorithms.

56. <https://www.pooldata.io/our-platform>

57. <https://agdatahub.eu/en/api-agro/>

58. <https://www.nxtport.com/en/market/our-marketplace/marketplace>

4.5.3 Examples

- **API-AGRO**: is an agricultural data marketplace that coordinates the flow of data between operators from the private and public sector to improve the performance of agricultural producers and their value chains. In 2014, the French Ministry of Agriculture funded a research project that later turned into a company that manages the API-AGRO platform. API-AGRO acts as a neutral intermediary, i.e., it does not monetise data but mediates the connection between data holders and data users.
- **Nxtport** marketplace: provides infrastructure (technical and legal) which enables: (i) the sharing of information in a secure environment, keeping full control over data to supply chain partners of ports worldwide, and (ii) access to relevant information by all stakeholders — including shipping operators, service providers, port authorities, government agencies.
- **CARUSO** is a platform offering harmonized multi-brand in-vehicle data from different vehicle manufacturers. CARUSO is co-owned by multiple companies in the automotive industry. It positions its offering as a neutral intermediary that obtains data from original equipment manufacturers, harmonises it and makes it available to all players in the mobility sector. Some examples of the latter include insurance companies, car sharing platforms and apps allowing drivers to find and book electric charging stations.

4.5.4 Business model considerations

The main challenge faced by data marketplaces in terms of business model design consists in building an attractive value proposition. By definition, a data marketplace does not offer data analytics services (i.e., services based on the analysis of the data), instead it facilitates data exchanges. This implies that its value proposition relies on reducing search and transaction costs (e.g., by providing services such as tooling for facilitating data exchanges or standardised technical solutions and contracts); otherwise, transactions would take place outside the marketplace. In some cases, the value proposition can include two additional pillars: aggregating data and/or increasing data quality.

In some industries, search and transaction costs are not high enough, or the number of potential transactions is not large enough to justify the need for a data marketplace. For example, in the aviation industry, the duopoly between the two main producers makes them the natural brokers of hub-and-spoke data exchanges with their suppliers and airlines. As a result, Airbus launched its data-sharing platform, Skywise, for airlines to share operational and flight data with Airbus in exchange for operational and revenue efficiencies (i.e., in exchange for a data-driven service).

In contrast, for a pure data marketplace to succeed in building an attractive value proposition, the organisations seeking to exchange data have to (a) be atomised; (b) require a sufficiently high volume of transactions; and (c) face high search and transaction costs (e.g., lack of awareness about other actors and their data, and lack of standardisation, high liability risks involved in the exchange of data, etc.). In such a case, a data marketplace can emerge as an

attractive value proposition. An example of such a case is the connected vehicles ecosystem, where third party (i.e., not linked to manufacturers) data marketplaces such as CARUSO⁵⁹ or Otonomo⁶⁰ have successfully emerged in the early stage of the evolution of this ecosystem.

Consequently, as noted by the OECD (2019), *despite the growth of data intermediaries, there is no single data marketplace where organisations and individuals can sell or exchange data directly with each other and many fail to scale. Microsoft's DataMarket and Data Services solutions, for instance, were integrated in Microsoft's cloud computing platform (Microsoft Azure). The uptake of both services has been, however, not as expected, forcing Microsoft to discontinue both services as of March 2017* (OECD, 2019, p. 39).

4.6 Data sharing pools

4.6.1 Introduction

Data sharing pools are alliances among data holders that share data with the aim of improving their assets (data products, processes and services) by exploiting the complementarities of the pooled data. These alliances develop around a shared purpose, context or application, and are intended to benefit all of their participants. In the context of this report, the data sharing pool concept does not link to any technical architecture style but focuses on the efforts needed for cooperation and coordination — at technical, legal and organisational levels — among the different actors. The nature of this coordination will depend on the type of stakeholders, their relationships, and the overall purpose of the pool. Data interme-

59. <https://www.caruso-dataplace.com/>

60. <https://otonomo.io>

diation mechanisms in data sharing pools should ensure a fair distribution of value within the pool without disrupting fair competition rules of the market.

Stimulating data sharing to foster innovation is a priority for European policymakers, as recognised in several communications (Data strategy,⁶¹ Industrial strategy,⁶² SME strategy⁶³). Promoting data sharing pools is one strategy contributing to achieving this policy goal through enhanced cooperation among stakeholders. Cooperation, among data holders, data subjects and/or data service providers, is in fact an important precondition for innovation (Mattioli, 2017).

4.6.2 How they work

Data sharing pools aim to attract actors to cooperate through sharing data with a shared purpose (e.g., advance research on a specific disease), context (e.g., improve efficiency of the energy supply) or application (e.g., mobility as a service in a city).

The data intermediation role within data sharing pools can exist at two levels, as a centralised pool orchestrator or as a service provider partner within the association. A pool orchestrator can manage the access and uses of data held in the pool by defining specific governance mechanisms through contractual means. The overall objective of these mechanisms is to instil trust among the pool participants and reduce contextual barriers (technical or co-ordinational). For this, the pool manager must oversee data access and their use by participants and ensure the compliance with agreed terms.

Data sharing pools complying with inclusive data governance principles would function

in a way to reduce power asymmetries among participants and to fairly distribute value generated from data among them. Moreover, the products and services created thanks to the pool would not impact negatively on competition forces on related markets. An example of a data sharing pool with an inclusive data governance approach would be an alliance of actors with complementary data to improve research on a ‘rare disease’ and the development of effective drugs. The pool would ensure that (i) the governance of access to these data and the policies for their use are defined in a participatory manner, (ii) the pool is accessible to a broad set of actors (e.g., researchers and SMEs), (iii) the pool allows for a fair share of value distribution within the pool (power asymmetries are mitigated), and (iv) the pool does not abuse market power by setting unaffordable prices for the products developed.

4.6.3 Examples

- **UK Biobank**⁶⁴ is a large-scale biomedical database that aims at enabling new scientific discoveries to improve public health. It provides accredited researchers access to medical and genetic data from half a million volunteer participants to improve understanding of the prevention, diagnosis and treatment of a wide range of serious and life-threatening illnesses.
- **INSIGHT**, The Health Data Research Hub for Eye Health,⁶⁵ aims at making routinely collected eye data available for research. All data related to patients receiving routine care at two UK National Health Service (NHS) foundation trusts — Moorfields Eye Hospital and University Hospitals Bir-

61. <https://digital-strategy.ec.europa.eu/en/policies/strategy-data>

62. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-industrial-strategy_en

63. <https://digital-skills-jobs.europa.eu/en/actions/european-initiatives/sme-strategy-sustainable-and-digital-europe>

64. <https://www.ukbiobank.ac.uk/>

65. <https://www.insight.hdrhub.org/>

mingham — are collected in INSIGHT, and patients can decide at any time to opt-out. Interested researchers can apply for access to INSIGHT datasets by submitting a request and a proposal. The business model applied by this initiative is a cost-recovery one, meaning that researchers will be charged only the cost of administering and processing the data request.

- The **Pistoia Alliance**⁶⁶ is a not-for-profit organisation created in 2009 by a group of pharmaceutical companies (AstraZeneca, GSK, Novartis and Pfizer) with the aim of overcoming common obstacles to innovation and of transforming R&D. It is currently composed of more than 100 member companies of different sizes (large, medium or individuals and start-ups) as well as government, academia and not-for-profit entities. The Pistoia Alliance is structured under various projects to which researchers and funders may decide to contribute.
- **BRUcloud**⁶⁷ is an open data sharing platform for the cargo community of Brussels airport, allowing the different stakeholders in the air cargo supply chain to work in an integrated way. Data are stored only once, centrally, and single cargo companies can use existing applications and start exchanging information with other stakeholders. BRUcloud is based on the data-sharing technology developed by Nallian,⁶⁸ and allows the data holders to remain in control of their data and to define which data-field is shared with whom and for which purpose.

66. <https://www.pistoiaalliance.org/>

67. <https://brucloud.com/>

68. <https://nallian.com/>

4.6.4 Business model considerations

As the examples above illustrate, data sharing pools are typically constituted by organisations that have complementary datasets. By pooling their data, they all benefit from economies of scale and/or scope in data aggregation. The value proposition of data sharing pools therefore consists precisely in exploiting these economies so that all the members can grasp them by using the pooled data.

Operating a data pool entails a series of fixed (e.g., administrative costs) and variable (e.g., cloud services, legal costs to sign in new members, etc.) costs. Because data pools' value proposition consists in providing access to a shared resource, the revenue streams necessary to cover the costs usually take the form of subscription fees by the members. These fees can differ between members based on several criteria that are specific to the objectives of the data pool. For example, some pools might simply have a fixed fee that is the same for all members. Other pools might decide to charge a variable fee that depends on the level of use of the pooled data. Others might reduce the fees to the members that contribute the most or the highest-quality data to the pool.

In some cases, some types of members (e.g., not-for-profit organisations, academic institutions) might pay less or nothing, and the other members subsidise them. This can be the case if the data pool has non-economic objectives that can exist alongside economic ones, or if the use of the data by certain organisations is economically beneficial to other members of the data pool. Alternatively, some organisations can be subsidised to create incentives for them to contribute data to the pool (Carballa-Smichowski et al., 2021).

5 Concluding remarks

This report is the first outcome from a series of science for policy activities that the JRC is conducting to support the European Strategy for Data (European Commission, 2020b), and in particular the implementation of the Data Governance Act (European Commission, 2022). It presents the results of a desk research study, strengthened by the feedback collected during an online expert workshop with academics and specialists affiliated to other institutions.

The goal of the report is to provide both a description and an analysis of the landscape of data intermediaries, especially those which may have the potential to foster more **inclusive forms of data governance**. The report strives to create a bridge between the wider landscape of data intermediaries and the concept of inclusive data governance, as well as the current EU regulatory framework, in particular the Data Governance Act (which entered into force on 23 June 2022 and will be applicable from 24 September 2023). The concept of inclusive data governance indicates a (possible) direction for the governance of data in Europe and beyond: its principles and underpinnings are based on current debates in the policy and academic literature, as well as on the founding values of the EU, which stand at the centre of the EU Charter of Fundamental Rights and the European approach to the digital transformation.

The report is intended to serve as an informative resource for stakeholders involved in different roles and activities relating to data intermediaries. It aims to provide conceptual clarity on the broad and multifaceted

landscape of data intermediaries that may promote alternative and fairer approaches to data governance and use, especially compared to the practices of large online platforms. It also examines the relationship of such intermediaries to the DGA. Furthermore, it provides an overview of their business models and addresses the key issue of the economic sustainability of data intermediaries, which is a prerequisite for their successful diffusion in Europe, among others. In particular, the success of a data intermediary requires building trust and promoting the willingness of both data holders and data subjects to adopt its services.

The review of data intermediaries in this report draws from current policy and academic literature. The overview of the **six types of data intermediaries** that, at least to a certain extent, can be said to promote more inclusive forms of data governance, highlights the fragmentation and heterogeneity of the field. Data intermediaries that are currently most often addressed by practitioners and experts range from individualistic and business-oriented types to more collective and inclusive models that support greater engagement in data governance by communities and individual data subjects. Some data intermediaries focus the intermediation on technical solutions and infrastructures (e.g., PIMs and data marketplaces), while others use legal constructs (e.g., data trusts), or other collective governance mechanisms (e.g., data cooperatives). Furthermore, while certain types aim at facilitating economic transactions between data holders and users, others mainly seek to produce collective benefits or public value.

Not all the data intermediaries examined in the report can be said to equally foster inclusive data governance. As raised by several experts during a feedback workshop on this report, market-oriented or highly individualistic approaches, aimed at increasing access or control over data by economic entities or by data subjects, may only address in a limited manner the structural power asymmetries of the current data landscape. For instance, as claimed by Janssen et al. in relation to data intermediaries for personal data, ‘consent-based grounds for processing data’ — without other policy interventions and governance mechanisms — risk having a limited impact ‘in [their] capacity to protect users in digital ecosystems because [they] put more responsibility on single individuals to address issues which for a large part are beyond their control’ (2020: 18-19).

Furthermore, as already explained in previous sections of the report, not all six types of data intermediaries that are presented here will necessarily encompass initiatives that fulfil the requirements of the DGA for data intermediation services providers (DISPs). The DGA regulates DISPs with the specific goal to increase trust in neutral third parties that enable economic (data) transactions between data subjects/holders and data users, thus addressing only to a certain extent the aims of inclusive data governance, but not completely or comprehensively overlapping with this definition.

Drawing on the review of the landscape of data intermediaries that is presented in the report, the following **main take-aways** can be derived regarding the **obstacles faced by data intermediaries**:

- From the current scenario, it is not clear which **business models** can guarantee the economic sustainability of the different types of data intermediaries. Data intermediaries that

comply with the requirements set out in Chapter III of the DGA must adopt a business model that does not depend on developing data-driven services from the data that are shared through them. If a data intermediary aims to conduct data analytics and derive insights and data-driven services from such data, this activity should take place through a separate legal entity. Business models that abide by the DGA requirements might present challenges in terms of economic sustainability and outcomes. For instance, if certain data intermediaries need to maintain high membership fees to sustain themselves, this could relegate them to a niche service, accessible only to well-resourced individuals or larger companies.

- To stimulate the **demand for data intermediation services**, the advantages must be apparent and appreciated. Potential users of data intermediaries need a sufficient level of **awareness, knowledge and expertise in data and digital matters** to understand the implications of data processing and transfer, and the benefits that may be offered in this regard by trustworthy DISPs. The demand for data intermediaries could thus potentially grow in the future through greater awareness about data asymmetries and the existence of trustworthy DISPs that comply with DGA requirements. However, other aspects of the demand should also be considered for a successful uptake of these new services. These include, on the one hand, recognising that digital divides and an absence of technological literacy persist among certain segments of the EU population, and on the other hand, recognising that certain issues relating to

data governance are complex to understand even for expert users, due to a certain level of technological or legal sophistication. Current adoption of data intermediaries by individuals seems limited and participation in this space often tends to be ‘biased’ towards the more privileged actors. Several questions remain open, such as: how do we deal with the fact that there will be a significant number of people that may not be aware of the existence of data intermediaries built to empower them and the communities they are part of? On a more structural and societal level, it is also important to consider how people could be more meaningfully involved in decision-making and control of their data, including in the context of data intermediaries.

- Uncertainty about the **neutrality requirement** for DISPs in the DGA regulation. The reference to neutrality in the DGA does not imply that providers of data intermediation services cannot derive any economic value from their intermediation activities. They are allowed to make revenues, but these should not originate from an analysis of the data they intermeditate, otherwise a conflict of interests might arise. A related issue concerns whether DISPs (regulated by the DGA) are allowed to aggregate and analyse data. This activity is not forbidden by the DGA requirements, but DISPs are not allowed to draw from such data aggregation and analysis in order to build services that will provide them with a revenue stream, unless this is carried out through a separate legal entity — hence, building data-driven services upon intermediated data cannot be their main business model.

This report is intended to provide only an overview of the landscape of data intermediaries, and it inevitably leaves some areas unexplored. Arising from some of the gaps in this report, we identify possible directions **for future science for policy research** in this field. We suggest that the following aspects could be investigated further:

- The novelty of data intermediation services providers and their impact on data governance calls for a deeper understanding of the **links and connections with other EU data policy instruments**. Future research in this field could explore: the potential role of data intermediaries in cross-border data flows and digital trade; the potential role of data intermediaries as neutral facilitators in the implementation of data access rights established by the Data Act; and the potential role of data intermediaries in common European data spaces. On the latter, future research can explore how data intermediaries can be enablers and facilitators of data reuse in a common European data space, serving as matchmakers between the demand and supply of data, and could also investigate how European data spaces might serve as controlled environments in which to develop and adopt interoperability standards for data intermediaries, eventually supporting their wider uptake.
- It is also important to consider the **needs of developers and entrepreneurs** that are active in the field of data intermediaries and related entities, such as data altruism organisations. Empirical research could inform policy makers on how to provide accessible and user-friendly information about the issues at stake in setting up data intermediaries, as

well as guidance concerning the requirements of current regulations (especially the DGA). Such guidance could cover business models and legal aspects, as well as the purposes and governance features of data intermediaries that promote more inclusive forms of data governance. Additionally, this line of work could address technical and infrastructural needs of data intermediaries, which have not been dealt with in this report but nevertheless play an important role in the establishment and operation of data intermediaries.

- **The supply and demand side of data intermediaries.**

- Research could explore the **incentives for setting up data intermediaries**, conditions for ensuring sustainability, the expected market size, the minimum scale for their effectiveness and efficiency, and the maximum scale to avoid market and power concentration. This would also include identifying and exploring the various factors that could support the building of trust among the different actors involved in data intermediation: data subjects, data holders and data users.
- Research could explore the operational aspects of data intermediaries, moving beyond the current focus on the legal infrastructures and governance frameworks (especially for certain types, such as data trusts). More attention could be placed on human resources, skillsets, **professional practices** and roles that are needed for the establishment and operation of data intermediaries. For example, researching ‘data stewardship’ as

a profession, as already pointed to by the emerging literature about data stewards.

- Research could explore to what extent **users are ready** for data intermediaries. A future important area of research on the demand side concerns how to meaningfully communicate and manage the risks and benefits involved in sharing data (Janssen et al., 2020). Such communication and protection/risk management mechanisms concern both data intermediaries directly and other initiatives. The latter ranges from data literacy interventions to how platforms adopt tools to improve transparency and communicate with their users about their data handling practices.
- Data intermediary incentives can be controversial (e.g., financial incentives for sharing health data), as they might put those who already experience vulnerability in an even more vulnerable position. Research could explore what **mitigation measures** can be put in place and how a **trustworthy design** can be supported so that data intermediaries do not become subject to the same motivations, incentives and ills of the current main players in the digital economy.
- Once the DGA has been applicable for a sufficient period, the JRC could investigate through empirical analysis the **trend of emerging data intermediaries**, engaging again with the experts that were involved in the workshop for consolidating this report.

References

- Abiteboul, S., B. André, and D. Kaplan, 'Managing Your Digital Life', *Communications of the ACM*, Vol. 58, No. 5, April 23, 2015, pp. 32–35. <https://doi.org/10.1145/2670528>
- Abraham, R., J. Schneider, and J. vom Brocke, 'Data Governance: A Conceptual Framework, Structured Review, and Research Agenda', *International Journal of Information Management*, Vol. 49, December 1, 2019, pp. 424–438. <https://doi.org/10.1016/j.ijinfomgt.2019.07.008>
- Ada Lovelace Institute, *Participatory data stewardship. A framework for involving people in the use of data*, 2021. <https://www.adalovelaceinstitute.org/report/participatory-data-stewardship/>
- Ada Lovelace Institute, *Rethinking data and rebalancing digital power*, 2022. <https://www.adalovelaceinstitute.org/report/rethinking-data/>
- Allard, T., N. Anciaux, L. Bouganim, Y. Guo, L. Le Folgoc, B. Nguyen, P. Pucheral, I. Ray, I. Ray, and S. Yin, 'Secure Personal Data Servers: A Vision Paper', *Proceedings of the VLDB Endowment*, Vol. 3, No. 1–2, September 2010, pp. 25–35. <https://doi.org/10.14778/1920841.1920850>
- Anciaux, N., P. Bonnet, L. Bouganim, B. Nguyen, P. Pucheral, I. Sandu Popa, and G. Scerri, 'Personal Data Management Systems: The Security and Functionality Standpoint', *Information Systems*, Vol. 80, February 2019, pp. 13–35. <https://doi.org/10.1016/j.is.2018.09.002>
- Barreau, D.K., 'Context as a Factor in Personal Information Management Systems', *Journal of the American Society for Information Science*, Vol. 46, No. 5, 1995. [https://doi.org/10.1002/%28SICI%291097-4571%28199506%2946%3A5%3C327%3A%3AAID-ASI4%3E3.O.CO%3B2-C](https://asistdl.onlinelibrary.wiley.com/doi/epdf/10.1002/%28SICI%291097-4571%28199506%2946%3A5%3C327%3A%3AAID-ASI4%3E3.O.CO%3B2-C)
- Batini, C., and M. Scannapieco, *Data Quality: Concepts, Methodologies and Techniques, Data-Centric Systems and Applications*, Springer, Berlin, 2006.
- Beaulieu, A., and S. Leonelli, *Data Quality: Concepts, Methodologies and Techniques*, Sage, London, 2022.
- Blasimme, A., E. Vayena, and E. Hafen, 'Democratizing Health Research Through Data Cooperatives', *Philosophy & Technology*, Vol. 31, No. 3, September 2018, pp. 473–479. <https://doi.org/10.1007/s13347-018-0320-8>
- Bodó, B., K. Irion, H. Janssen, and A. Giannopoulou, 'Personal Data Ordering in Context: The Interaction of Meso-Level Data Governance Regimes with Macro Frameworks', *Internet Policy Review*, Vol. 10, No. 3, September 30, 2021. <https://doi.org/10.14763/2021.3.1581>
- Borgogno, O., and G. Colangelo, 'Data Sharing and Interoperability: Fostering Innovation and Competition through APIs', *Computer Law & Security Review*, Vol. 35, No. 5, October.
- Borkin, S., *Platform Co-Operatives. Solving the Capital Conundrum*, Nesta and Co-operatives UK, 2019. <https://doi.org/10.1016/j.clsr.2019.03.008>
- Carballa Smichowski, B., 'Alternative Data Governance Models: Moving Beyond One-Size-Fits-All Solutions', *Intereconomics*, Vol. 54, No. 4, July 2019, pp. 222–227. <https://doi.org/10.1007/s10272-019-0828-x>

Carballa Smichowski, B., N. Duch Brown, and B. Martens, 'To Pool or to Pull Back? An Economic Analysis of Health Data Pooling', JRC Working Papers on Digital Economy. Joint Research Centre, June 2021.

Carballa Smichowski, B., N. Duch, S. Höcük, P. Kumar, B. Martens, J. Mulder, and P. Prufer, *Economies of Scope in Data Aggregation: Evidence from Health Data*, 2022. https://www.researchgate.net/publication/365276562_Economies_of_scope_in_data_aggregation_evidence_from_health_data

Centre for Data Ethics and Innovation, *Unlocking the Value of Data: Exploring the Role of Data Intermediaries*, 2021. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4338447

Charisi, V., S. Chaudron, R. Di Gioia, R. Vuorikari, M. Escobar Planas, J.I. Sanchez Martin, and E. Gomez Gutierrez, *Artificial Intelligence and the Rights of the Child: Towards an Integrated Agenda for Research and Policy*, EUR 31048 EN, Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92-76-51837-2, doi:10.2760/012329, JRC127564.

Cities Coalition for Digital Rights, and UN HABITAT for a better urban future, 'Digital Rights Governance Framework. Concept Draft Open for Feedback', *Cities for Digital Rights*, 2022. <https://citiesfordigitalrights.org/first-draft-digital-rights-governance-framework-out-and-open-feedback>

Coriat, B., *Quels Modèles Économiques Pour Les Communs ? Le Cas Des Plateformes Substantives* Programmes de Recherche TAPAS et EnCommuns WP EnCommuns N° 41, 2021. https://drive.google.com/file/d/1DhMhG6fz9-RqOaFdJkKfmZN-9j4btt_Xl/view?usp=sharing

Craglia, M., H.J. Scholten, M. Micheli, J. Hradec, I. Calzada, S. Luitjens, M. Ponti, and J.

Boter, *Digitranscope: The governance of digitally-transformed society*, EUR 30590 EN, Publications Office of the European Union, Luxembourg, 2021, ISBN 978-92-76-30229-2, doi:10.2760/503546, JRC123362.

Ctrl-Shift, *Personal Information Management Services: An Analysis of an Emerging Market*, 2014. <https://media.nesta.org.uk/documents/personal-information-management-services.pdf>

Data Administration Management Association, *The DAMA Guide to the Data Management Body of Knowledge: DAMA-DMBOK Guide*, Edited by M. Mosley, M. Brackett, S. Earley, and D. Henderson, First edition., Technics Publications, LLC, Bradley Beach, NJ, 2010.

Data Spaces Support Centre (DSSC), *Starter Kit for Data Space Designers*, 2022.

de Montjoye, Y.-A., E. Shmueli, S.S. Wang, and A.S. Pentland, 'OpenPDS: Protecting the Privacy of Metadata through SafeAnswers', Edited by T. Preis, *PLoS ONE*, Vol. 9, No. 7, July 9, 2014. <https://doi.org/10.1371/journal.pone.0098790>

Delacroix, S., and N.D. Lawrence, 'Bottom-up Data Trusts: Disturbing the 'One Size Fits All' Approach to Data Governance', *International Data Privacy Law*, October 1, 2019. <https://doi.org/10.1093/idpl/ipz014>

Element AI, and NESTA, *Data Trusts: A New Tool for Data Governance*, 2019. https://hello.elementai.com/rs/024-OAQ-547/images/Data_Trusts_EN_201914.pdf

EPRS (European Parliamentary Research Service), *Governing Data and Artificial Intelligence for All. Models for Sustainable and Just Data Governance*, STOA | Panel for the Future of Science and Technology, 2022. [https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU\(2022\)729533](https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2022)729533)

European Commission, REGULATION (EU) 2016/679 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 and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), 2016.

European Commission, *Commission Staff Working Document. Impact Assessment Report. Accompanying the Document Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on European Data Governance (Data Governance Act)*, 2020a. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020SC0295>

European Commission, *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. A European Strategy for Data (COM(2020)66)*, 2020b. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0066>

European Commission, *REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on European Data Governance and Amending Regulation (EU) 2018/1724 (Data Governance Act)*, 2022. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32022R0868>

Farrell, E., Minghini, M., Kotsev, A., Soler Garrido, J., Tapsall, B., Micheli, M., Posada Sanchez, M., Signorelli, S., Tartaro, A., Bernal Cereceda, J., Vespe, M., Di Leo, M., Carballa Smichowski, B., Smith, R., Schade, S., Pogorzelska, K., Gabrielli, L. and De Marchi, D., *European Data Spaces - Scientific Insights into Data Sharing and Utilisation at Scale*, EUR 31499 EN, Publications Office of the European Union, Luxembourg, 2023, ISBN 978-92-68-03166-7, doi:10.2760/301609, JRC129900.

Floridi, L., 'Big Data and Information Quality', in L. Floridi and P. Illari (eds.), *The Philosophy of Information Quality*, Springer International Publishing, Cham, 2014, pp. 303–315.

Giannopoulou, A., J. Ausloos, S. Delacroix, and H. Janssen, 'Intermediating Data Rights Exercises: The Role of Legal Mandates', *International Data Privacy Law*, Vol. 12, No. 4, December 7, 2022, pp. 316–331. <https://doi.org/10.1093/idpl/ipac017>

Global Partnership on Artificial Intelligence (GPAI), Aapti Institute, and Open Data Institute, *Enabling Data Sharing for Social Benefit through Data Trusts*, 2021. <https://gpai.ai/projects/data-governance/data-trusts/enabling-data-sharing-for-social-benefit-through-data-trusts.pdf>

Hardinges, J., and J.R. Keller, *What are data institutions and why are they important?*, 2021. <https://www.theodi.org/article/what-are-data-institutions-and-why-are-they-important/>

Ho, C., and T. Chuangt, 'Governance of Communal Data Sharing', by A. Daly, K. Devitt, and M. Mann, *Good Data*, Institute of Network Cultures, Amsterdam, NL, 2019.

Janssen, H., J. Cobbe, C. Norval, and J. Singh, 'Decentralised Data Processing: Personal Data Stores and the GDPR', *International Data Privacy Law*, Vol. 10, No. 4, November 2020, pp 356–384. <https://doi.org/10.1093/idpl/ipaa016>

Janssen, H., and J. Singh, 'Data Intermediary', *Internet Policy Review*, Vol. 11, No. 1, March 30, 2022a. <https://doi.org/10.14763/2022.1.1644>

Janssen, H., and J. Singh, 'Personal Information Management Systems', *Internet Policy Review*, Vol. 11, No. 2, April 11, 2022b. <https://doi.org/10.14763/2022.2.1659>

Kooiman, J., *Governing as Governance*, SAGE Publications Ltd, London, 2003. <https://doi.org/10.4135/9781446215012>

Koutroumpis, P., A. Leiponen, and L.D.W. Thomas, *The (Unfulfilled) Potential of Data Marketplaces*, ETLA Working Papers, The Research Institute of the Finnish Economy (ETLA), September 29, 2017. <https://www.etla.fi/wp-content/uploads/ETLA-Working-Papers-53.pdf>

Lähteenoja, V., 'What Are "Personal Data Spaces"?', *Companion Proceedings of the ACM Web Conference 2023*, ACM, Austin TX USA, 2023, pp. 1458–1461. <https://doi.org/10.1145/3543873.3587656>

Lähteenoja, V., and S. Sepp, *State of MyData 2021*, MyData Global, 2021. https://www.mydata.org/wp-content/uploads/2022/07/State-of-MyData-2021_reduced-2.pdf

Langford, Joss, Antti 'Jogi' Poikola, Wil Jansen, Viivi Lähteenoja, and Marlies Rikken, eds., *Understanding MyData Operators*, MyData Global, 2022. <https://www.mydata.org/wp-content/uploads/2022/07/Understanding-MyData-Operators-2022-1.pdf>

Lanier, J., and G.E. Weyl, 'A Blueprint for a Better Digital Society', *Harvard Business Review*, September 26, 2018. <https://hbr.org/2018/09/a-blueprint-for-a-better-digital-society>

Lehtiniemi, T., 'Personal Data Spaces: An Intervention in Surveillance Capitalism?', *Surveillance & Society*, Vol. 15, No. 5, December 5, 2017, pp. 626–639. <https://doi.org/10.24908/ss.v15i5.6424>

Liu, J., 'Social Data Governance: Towards a Definition and Model', *Big Data & Society*, Vol. 9, No. 2, July 2022. <https://doi.org/10.1177/20539517221111352>

Madiega, T., *Digital Sovereignty for Europe*, EPRS Ideas Paper. Towards a More Resilient EU, EPRS | European Parliamentary Research Service, 2020. [https://www.europarl.europa.eu/RegData/etudes/BRIE/2020/651992/EPRS_BRI\(2020\)651992_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2020/651992/EPRS_BRI(2020)651992_EN.pdf)

Mannan, M., J. Wong, and E. Bietti, 'Data Cooperatives in Europe: A Preliminary Investigation', *Network Industries Quarterly*, Vol. 24, No. 3, 2022. <https://www.network-industries.org/wp-content/uploads/2022/10/Data-Cooperatives-in-Europe.pdf>

Mattioli, M., 'The Data-Pooling Problem', *Berkeley Technology Law Journal*, Vol. 32, No. 1, 2017, pp. 179–236.

Meijer, A., and S. Potjer, 'Citizen-Generated Open Data: An Explorative Analysis of 25 Cases', *Government Information Quarterly*, Vol. 35, No. 4, October 1, 2018, pp. 613–621. <https://doi.org/10.1016/j.giq.2018.10.004>

Micheli, M., M. Ponti, M. Craglia, and A. Berti Suman, 'Emerging Models of Data Governance in the Age of Datafication', *Big Data & Society*, Vol. 7, No. 2, July 1, 2020. <https://doi.org/10.1177/2053951720948087>

Mozilla Insights, J. van Geuns, and A. Brădușescu, *Shifting Power Through Data Governance*, Mozilla Foundation, September 16, 2020. <https://assets.mofoprod.net/network/documents/ShiftingPower.pdf>

Mydex, *Achieving Transformation at Scale*, 2021. <https://mydex.org/resources/papers/AchievingTransformationAtScale/>

OECD, *Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-Use across Societies*, OECD Publishing, Paris, 2019. <https://doi.org/10.1787/276aa-ca8-en>

Olson, J.E., *Data Quality: The Accuracy Dimension*, Morgan Kaufmann, San Francisco, 2003.

Open Data Institute, *What is a data trust?*, July 10, 2018. <https://www.theodi.org/article/what-is-a-data-trust/>

Ponti, M., and M. Craglia, *Citizen-Generated Data for Public Policy*, European Commission, Ispra, 2020.

Posner, E.A., and G.E. Weyl, *Radical Markets: Uprooting Capitalism and Democracy for a Just Society*, Princeton University Press, Princeton, NJ, 2018.

Prainsack, B., and N. Forgó, 'Why Paying Individual People for Their Health Data Is a Bad Idea', *Nature Medicine*, Vol. 28, No. 10, October 2022, pp. 1989–1991. <https://doi.org/10.1038/s41591-022-01955-4>

RadicalxChange Foundation Ltd, 'The Data Freedom Act', 2020. <https://www.radicalxchange.org/media/papers/data-freedom-act.pdf>

Reed, C., B. Solicitors, and P. Masons, 'Data Trusts: Legal and Governance Considerations', *Pinsent Masons*, 2019. <http://theodi.org/wp-content/uploads/2019/04/General-legal-report-on-data-trust.pdf>

Rosenbaum, S., 'Data Governance and Stewardship: Designing Data Stewardship Entities and Advancing Data Access: Data Governance and Stewardship', *Health Services Research*, Vol. 45, No. 5p2, October 2010, pp. 1442–1455. <https://doi.org/10.1111/j.1475-6773.2010.01140.x>

Sadowski, J., S. Viljoen, and M. Whittaker, 'Everyone Should Decide How Their Digital Data Are Used — Not Just Tech Companies', *Nature*, Vol. 595, No. 7866, July 2021, pp. 169–171. <https://doi.org/10.1038/d41586-021-01812-3>

Sandoval, M., 'Entrepreneurial Activism? Platform Cooperativism Between Subversion and Co-Optation', *Critical Sociology*. Vol 46, No 6, 2020, pp. 801-817. <https://doi.org/10.1177/0896920519870577>

Savona, M., 'Governance Models for Redistribution of Data Value', *VoxEU*, January 17, 2020. <https://cepr.org/voxeu/columns/governance-models-redistribution-data-value>

Scholz, T., *Platform Cooperativism: Challenging the Corporate Sharing Economy*, Rosa Luxemburg Stiftung, New York Office, 2016.

Smith, D., 'Governing Data and Data for Governance: The Everyday Practice of Indigenous Sovereignty', *Indigenous Data Sovereignty: Toward an Agenda*, ANU Press, 2016, pp. 117–135.

Spiekermann, M., 'Data Marketplaces: Trends and Monetisation of Data Goods', *Intereconomics*, Vol. 54, No. 4, July 2019, pp. 208–216.

Sutton, M., 'A Shareable Explainer: What Is a Platform Co-Op?', *Shareable*, 2016. <https://www.shareable.net/a-shareable-explainer-what-is-a-platform-co-op/>

Tengberg, J.C.F., 'Inter-Organizational Information Sharing of Customer Data in Retail', Diss. Massachusetts Institute of Technology, 2013. <https://dspace.mit.edu/handle/1721.1/80678>

UK AI Council, and Ada Lovelace Institute, *Exploring Legal Mechanisms for Data Stewardship, Data for the Public Good*, 2021. https://www.adalovelaceinstitute.org/wp-content/uploads/2021/03/Legal-mechanisms-for-data-stewardship_report_Ada_AI-Council-2.pdf

Urquhart, L., N. Sailaja, and D. McAuley, 'Realising the Right to Data Portability for the Domestic Internet of Things', *Personal and Ubiquitous Computing*, Vol. 22, No. 2, April 2018, pp. 317–332. <http://dx.doi.org/10.1007/s00779-017-1069-2>

Van Alstyne, M.W., G.G. Parker, and Choudary Sangeet Paul, '6 Reasons Platforms Fail', *Harvard Business Review*, March 31, 2016. <https://hbr.org/2016/03/6-reasons-platforms-fail>

Verhulst, S., 'Reimagining data responsibility: 10 new approaches toward a culture of trust in re-using data to address critical public needs', *Data & Policy*, Vol. 3, 2021. [doi:10.1017/dap.2021.4](https://doi.org/10.1017/dap.2021.4)

von Ditfurth, L., and G. Lienemann, 'The Data Governance Act: – Promoting or Restricting Data Intermediaries?', *Competition and Reg-*

ulation in *Network Industries*, Vol. 23, No. 4, December 2022, pp. 270–295. <https://doi.org/10.1177/17835917221141324>

Wang, R.Y., ‘A Product Perspective on Total Data Quality Management’, *Communications of the ACM*, Vol. 41, No. 2, February 1998, pp. 58–65.

World Economic Forum, *Advancing Digital Agency: The Power of Data Intermediaries*, February 15, 2022. <https://www.weforum.org/reports/advancing-digital-agency-the-power-of-data-intermediaries/>

Zingales, N., ‘Data Collaboratives, Competition Law and the Governance of EU Data Spaces’, *Research Handbook on the Law and Economics of Competition Enforcement*, 2022, pp. 8–49. <https://www.elgar-online.com/display/book/9781789903799/book-part-9781789903799-11.xml>

Zygmuntowski, J.J., L. Zoboli, and P.F. Nemitz, ‘Embedding European Values in Data Governance: A Case for Public Data Commons’, *Internet Policy Review*, Vol. 10, No. 3, September 30, 2021. <https://doi.org/10.14763/2021.3.1572>

List of abbreviations and definitions

AI	Artificial Intelligence	GPAI	Global Partnership on Artificial Intelligence
B2B	Business-to-Business	H2020	Horizon 2020
B2G	Business-to-Government	JRC	Joint Research Centre
CIC	Community Interest Company	NGOs	Non-Governmental Organisations
CNECT	Communications Networks, Content and Technology	NHS	UK National Health Service
DA	Data Act	ODI	Open Data Institute
DGA	Data Governance Act	OECD	Organisation for Economic Co-operation and Development
DMA	Digital Markets Act	PDMS	Personal Data Management Systems
DSA	Digital Services Act	PDS	Personal Data Store
DSP	Data Sharing Pool	PET	Privacy-Enhancing Technology
EC	European Commission	PIMS	Personal Information Management Systems
EDIB	European Data Innovation Board	R&D	Research and Development
EPRS	European Parliamentary Research Service	SME	Small and medium-sized enterprise
EU	European Union	UK	United Kingdom
G-Cloud	UK Government Digital Marketplace	UN	United Nations
GDPR	General Data Protection Regulation		
GISC	Grower's Information Service Cooperative		

List of boxes

Box 1. The Data Governance Act's provisions on data intermediation services providers (DISPs) in brief (Chapter III).

Box 2. The Data Governance Act's provisions on data altruism organisations (RDAOs) (Chapter IV).

Box 3. The concept of neutrality in the Data Governance Act.

Box 4. Correspondence between types of data intermediaries in Chapter III of the DGA and types described in this report.

Box 5. The Data Governance Act's provisions on data cooperatives.

Box 6. The Data Governance Act's provisions on data cooperatives and data altruism organisations.

List of figures

Figure 1. Overlaps between the different definitions of data intermediaries in this report.

List of tables

Table 1. DGA provisions for DISPs and RDAOs.

Table 2. Overview of the six types of data intermediaries covered in this report.



A decorative background at the top of the page featuring a light teal topographic map with various contour lines and shapes.

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct centres. You can find the address of the centre nearest you online (european-union.europa.eu/contact-eu/meet-us_en).

On the phone or in writing

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
- at the following standard number: +32 22999696,
- via the following form: european-union.europa.eu/contact-eu/write-us_en.

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website (european-union.europa.eu).

EU publications

You can view or order EU publications at op.europa.eu/en/publications. Multiple copies of free publications can be obtained by contacting Europe Direct or your local documentation centre (european-union.europa.eu/contact-eu/meet-us_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex (eur-lex.europa.eu).

Open data from the EU

The portal data.europa.eu provides access to open datasets from the EU institutions, bodies and agencies. These can be downloaded and reused for free, for both commercial and non-commercial purposes. The portal also provides access to a wealth of datasets from European countries.

Science for policy

The Joint Research Centre (JRC) provides independent, evidence-based knowledge and science, supporting EU policies to positively impact society



EU Science Hub

joint-research-centre.ec.europa.eu



@EU_ScienceHub



EU Science Hub - Joint Research Centre



EU Science, Research and Innovation



EU Science Hub



@eu_science



Publications Office
of the European Union