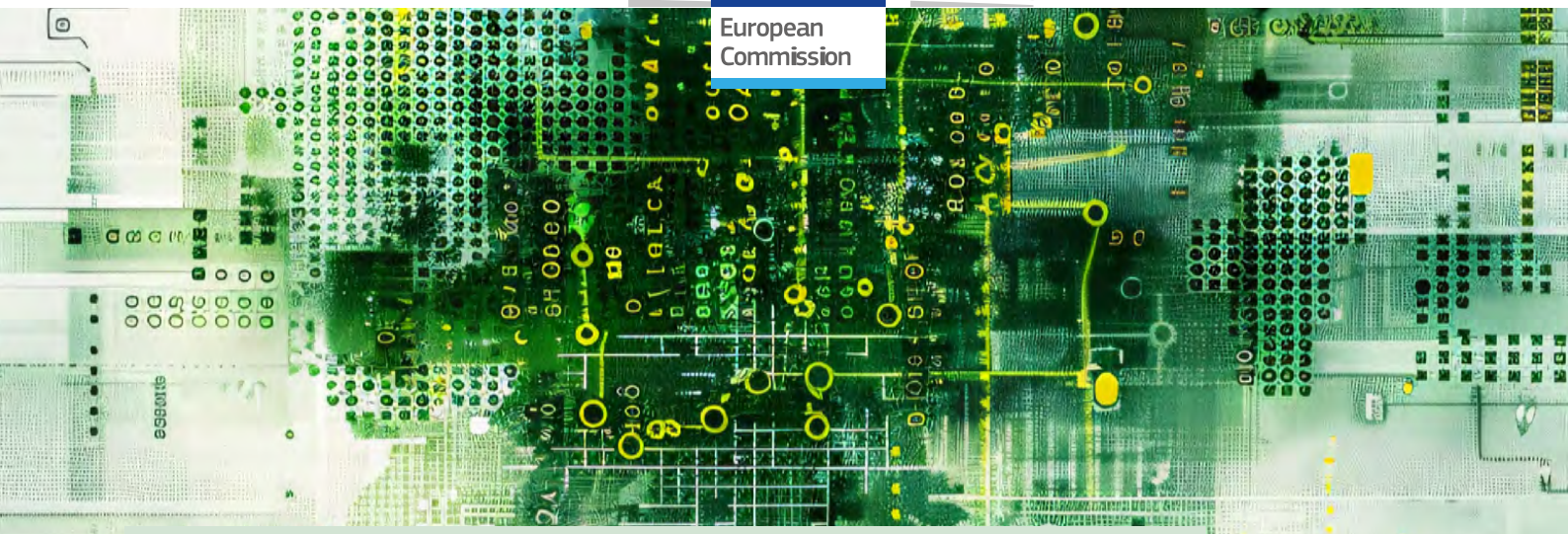
Unlocking Green Deal Data

Ponti, Marisa; Portela, Manuel; Pierri, Paola; Daly, Angela; Milan, Stefania; Kaukonen Lindholm, Riita

DOI:
10.2760/2422339
10.2760/0517622

Publication date:
2024

ISSN 1831-9424



Unlocking Green Deal Data: Innovative Approaches for Data Governance and Sharing in Europe

Authors: Ponti, M., Portela, M., Pierri, P., Daly, A., Milan, S., Kaukonen, R., Maccani, G., de Souza, S.P., Thabit, S.

2024

Joint Research Centre

EUR 40021

Contact information

Name: Alexander KOTSEV
Address: Via E. Fermi 2749, 21027 Ispra (VA) – Italy
Email: Alexander.KOTSEV@ec.europa.eu

Name: Sara THABIT GONZÁLEZ
Address: Via E. Fermi 2749, 21027 Ispra (VA) – Italy
Email: Sara.THABIT-GONZALEZ@ec.europa.eu

EU Science Hub
https://joint-research-centre.ec.europa.eu

JRC139026

EUR 40021

Print ISBN 978-92-68-21570-8 ISSN 1018-5593 doi:10.2760/2422339 KJ-01-24-011-EN-C
PDF ISBN 978-92-68-20584-6 ISSN 1831-9424 doi:10.2760/0517622 KJ-01-24-011-EN-N

Luxembourg: Publications Office of the European Union, 2024

How to cite this report: European Commission, Joint Research Centre, Ponti, M., Portela, M., Pierri, P., Daly, A., Milan, S., Kaukonen Lindholm, R., Maccani, G., Peter De Souza, S. and Thabit Gonzalez, S., Unlocking Green Deal Data: Innovative Approaches for Data Governance and Sharing in Europe, Thabit Gonzalez, S. and Maccani, G. (editors), Publications Office of the European Union, Luxembourg, 2024, https://data.europa.eu/doi/10.2760/0517622, JRC139026.

Layout: Carmen Capote de la Calle

Unlocking Green Deal Data: Innovative Approaches for Data Governance and Sharing in Europe

Authors: Ponti, M., Portela, M., Pierri, P., Daly, A., Milan, S., Kaukonen, R., Maccani, G., de Souza, S.P., Thabit, S.

2024

CONTENTS

ABSTRACT

Drawing upon the ambitious policy and legal framework outlined in the European Strategy for Data (2020) and the establishment of common European data spaces, this Science for Policy report explores innovative approaches for unlocking relevant data to achieve the objectives of the European Green Deal. The report focuses on the governance and sharing of Green Deal data, analysing a variety of topics related to the implementation of new regulatory instruments (i.e. the Data Governance Act and the Data Act) as well as the roles of various actors in the data ecosystem. It provides an overview of current incentives and disincentives for data sharing, and explores the existing landscape of data intermediaries and data altruism organisations. Additionally, it offers insights from a private sector perspective and outlines key data governance and sharing practices concerning citizen-generated data (CGD). Lastly, it provides a series of policy recommendations to support the ongoing revision of the INSPIRE Directive (2007), within the context of the common European Green Deal data space, and toward a more sustainable and fair data ecosystem.

FOREWORDS

Francesca Campolongo

Director, Digital Transformation and Data Directorate, European Commission Joint Research Centre (JRC)

Utilizing the potential of data and new technologies has become essential in addressing global environmental sustainability challenges. In the current European policy landscape, the EU Data Strategy (2020) has introduced a suite of regulatory and financial instruments with the potential to catalyse innovation and economic growth in Europe, aligning with President Ursula von der Leyen's 2025-2029 political priorities for a transformative 'data revolution' under a European Data Union Strategy. However, the challenge lies in the effective implementation of these instruments, ensuring they create a level playing field, particularly for smaller actors, and foster an environment that encourages regulatory innovation and simplification.

The Joint Research Centre (JRC) plays a crucial role in bridging the gap between ambitious policy initiatives and practical, actionable outcomes for navigating the complexities of the green and digital transition. In addition, the JRC extensive expertise in environmental data sharing and data governance, position it at the forefront of this transformation.

The present report provides a comprehensive analysis of the current environmental data governance landscape, emphasizing the importance of simplifying regulatory frameworks for broader access and participation in the data economy. The concept of 'Systemic Data Justice' introduced in the report underscores the need for equity, accountability, and fair representation to ensure the benefits of the data revolution are widely distributed without exacerbating social inequalities. It addresses the need for Europe to become more competitive while preserving social inclusion and tackling inequality, echoing the dual imperative highlighted in the Draghi report (2024).

This JRC's science for policy report is an important contribution to the ongoing discourse on Europe's digital and green transformation. It offers actionable insights for policymakers and stakeholders to address the challenges of the data economy, promoting an inclusive approach aligned with the European Green Deal's objectives. As Europe embraces this data revolution, the insights and recommendations from this report will guide the continent towards a technologically advanced and socially just future for all citizens.

From the JRC, we will persist in our efforts to support innovation, facilitate dialogue, and drive forward the policies that underpin our shared vision for a sustainable Europe.

Emil Andersen

Director, General Affairs, Knowledge & Resources, European Commission Directorate-General for Environment (DG ENV)

In an age when the twin digital and environmental transitions are reshaping societies, the European Commission's Directorate-General for Environment is committed to achieve the goals set out in the European Green Deal. This ambitious blueprint sets the stage for transforming the EU into a fair and prosperous society, with a modern, resource-efficient, and competitive economy. However, to achieve these goals, it is essential to harness digital innovations that empower us to monitor, understand, and manage environmental challenges with unprecedented precision and foresight.

This report, prepared by our EC Joint Research centre in close collaboration with our Directorate General, underscores the importance of sharing and governance mechanisms for environmental data in advancing the Green Deal transition, and highlights the opportunities presented by the forthcoming reviews of the current legal framework. It illustrates that effective implementation should prioritize simplicity and be undertaken in collaboration and dialogue with industry, social partners, and all stakeholders. By breaking down silos and fostering collaboration, we can not only expedite our response to environmental crises but also empower citizens, businesses, and communities to contribute to and benefit from the transition.

Furthermore, as we navigate these transformative pathways, the principle of a just transition leaving no one—and no place—behind remains at the core of our agenda. This report argues in favour of a data governance that prioritizes accessibility, inclusivity, and equity, thus enabling all sectors of society to participate in and drive the environmental discourse.

The path forward might be complex, but with a clear vision and collaborative effort, we can turn challenges into opportunities. This report serves as a call to action to step up, ensuring that our digital and environmental transitions lead to a future that is not only more sustainable and advanced, but also fairer for every European. We are committed to continue championing the use of digital innovations to tackle environmental challenges and ensure a socially fair transition. By working together with different stakeholders and embracing accessible data governance, we aim to create a greener and more sustainable future for all Europeans.

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to the following persons for their invaluable feedback and guidance throughout the entire process of this report: **Sotirios Kanellopoulos**, **Joeri Robbrecht**, **Rikke Nielsen** (European Commission, Directorate-General for Environment, DG ENV), **Alexander Kotsev**, **Sven Schade**, **Marco Minghini**, **Marina Micheli**, **Serena Signorelli**, **Eimear Farrell** (European Commission, Joint Research Centre), **Anna Berti Suman** (Libera Università Internazionale degli Studi Sociali - Luiss, Law School, Rome), and **Becky Kazansky** (University of Amsterdam).

Additionally, the authors would like to extend their honest thanks to the experts who have been involved and interviewed for chapter 4, for their time and unique contributions: **Claudio Agosti** (Hermes Center, AI Forensics, Reversing Works, Platform Auditor & Tech Researcher); **Heidrun Åm** (Professor in the Department of Sociology and Political Science at the Norwegian University for Science and Technology); **Paulina Behluli** (Programme Manager at The Data Tank); **Petter Falk** (Adjunct lecturer at Karlstad University); **Chrissy Martin Meier** (Policy Director at The Digital Impact Alliance); and **Barbara Prainsack** (Professor for Comparative Policy Analysis at the University Wien).

The authors are also grateful to **Sofia Kekempanou** (European Commission, Joint Research Centre) for her exceptional support in administrative matters, and to **Carmen Capote de la Calle** (European Commission, Joint Research Centre) for creating the layout of this report.

Authors [in order of chapters]

- **Marisa Ponti**, University of Gothenburg (Sweden).
- **Manuel Portela**, Universitat Pompeu Fabra (Spain)
- **Paola Pierri**, Bern University of Applied Sciences (Switzerland)
- **Angela Daly**, University of Dundee (Scotland/UK)
- **Stefania Milan**, University of Amsterdam (Netherlands) and European University Institute (Italy)
- **Riikka Kaukonen**, European Commission, Joint Research Centre (Italy), University of Helsinki (Finland)
- **Giovanni Maccani**, Ideas for Change (Spain)
- **Siddharth Peter de Souza**, independent law, data and society researcher.
- **Sara Thabit González**, European Commission, Joint Research Centre (Italy)

Editors:

- **Sara Thabit González**, European Commission, Joint Research Centre (Italy)
- **Giovanni Maccani**, Ideas for Change (Spain)

EXECUTIVE SUMMARY

Policy context

Unlocking the potential of data and innovative technologies is both a necessity and a political priority in the European context. It can play a crucial role in fulfilling the Commission's commitment to enable a **twin green and digital transition** to achieve the European Green Deal's political ambitions, including promoting energy efficiency, sustainable agriculture, biodiversity conservation, and a circular economy, to name just a few areas.

The 2024-2029 political guidelines of President-elect Ursula von der Leyen highlight these priorities and the fact that '**Europe needs a data revolution'**, emphasising the importance of accessing data in a trustworthy and secure manner to support productivity and societal innovations. Existing policy and legal frameworks outlined in the European Strategy for Data under von der Leyen's first Commission (2020) are pivotal to achieving these ambitions. In addition, in 2019 the Commission launched the GreenData4All initiative, with the goal of contributing to Europe's green and digital transformation by updating EU rules on environmental geospatial data and on public access to environmental information. As a result, the INSPIRE and the Public Access to Environmental Information Directives, which currently form the legal framework for environmental data sharing, are planned to be revised by 2025.

This report delves into the opportunities created by the instruments outlined in the European Strategy for Data and the revision of the existing legal frameworks on environmental data sharing, and provides policy recommendations for unlocking relevant data to achieve the objectives of the European Green Deal.

Key conclusions

The main conclusions build upon the concept of **'Systemic Data Justice'**, which emphasises equity, accountability, and fair representation to foster stronger connections between the supply and demand of data for a more effective and sustainable data economy.

Five policy recommendations outline a set of implications and actionable points for the revision of the INSPIRE Directive. The first point emphasises the need for a shift from public sector data to **public interest data**, which entails integrating a demand-driven perspective and exploring synergies between public and private interest data. The second recommendation underscores the importance of moving from a focus on data availability to **data use, participation, and agency** by addressing power imbalances and inclusion. Third, we highlight the necessity to move towards a **data ecosystem approach**, which fosters a data sharing and use economy that goes beyond a repository logic. The fourth recommendation emphasises the need to

foster increased trust as a key intended outcome, involving a focus on data quality, certification processes, and data literacy. Last, we advocate for **combining central governance rules with a decentralised approach** to data management, informed by use cases.

The relevance of these recommendations goes beyond Green Deal data only; they are instrumental to ensure that any data ecosystem is both inclusive and fair, addressing power imbalances and enabling more accessible and transparent data sharing and use practices.

Main findings

The main findings of the report focus on the opportunities and challenges that underpin the sharing and governance of Green Deal data, particularly concerning how their use can create positive impacts. Chapter 2 explains how existing legislation underpinning Green Deal data (i.e. Data Governance Act, Data Act, INSPIRE Directive, and Public Access to Environmental Information Directive), present both **incentives and disincentives for data sharing** among various players in the data economy. To be effective, regulatory incentives for data sharing and reuse require appropriate conditions related to trust, social capital, competition policies, and technical infrastructure. Chapter 3 analyses the **ecosystem of Green Deal data intermediation services**, and concludes that data providers, service providers, and data intermediaries are tightly interrelated in a highly complex ecosystem. It highlights the crucial role that technological enablers play in this context, and the importance of designing incentives to develop sustainable business models for data intermediation. Chapter 4 examines the **approaches and models of data altruism** and provides insights on the key actors and actions that are critical for this model to succeed. It highlights the need for adequate resources to cover the costs associated with data altruism, and the importance of incentives and use cases to operationalise this model. Chapter 5 examines the **participation of, and data contribution from businesses**, from a public good perspective. The chapter assesses opportunities and challenges for Business-to-Government (B2G), Business-to-Business (B2B) and Business-to-Consumer (B2C) data sharing, and exposes corporate social responsibility, accountability, and legal certainty as key drivers for Green Deal business data sharing. Chapter 6 revolves around the importance of **citizen-generated data (CGD) within the Green Deal data space**, reviewing best practices to foster and support the generation and utilisation of CGD. Additionally, Chapter 7 explains that implementing CARE (Collective Benefit, the Authority to Control, Responsibility and Ethics) principles in **data sharing with indigenous people and local communities**, is needed to ensure fair and sovereign approaches. The report closes with an overview of cross-cutting insights and policy recommendations in Chapter 8, providing **strategic advice and actionable points** for the revision of INSPIRE and unlocking the use of Green Deal data from a systemic data justice approach.

Related and future JRC work

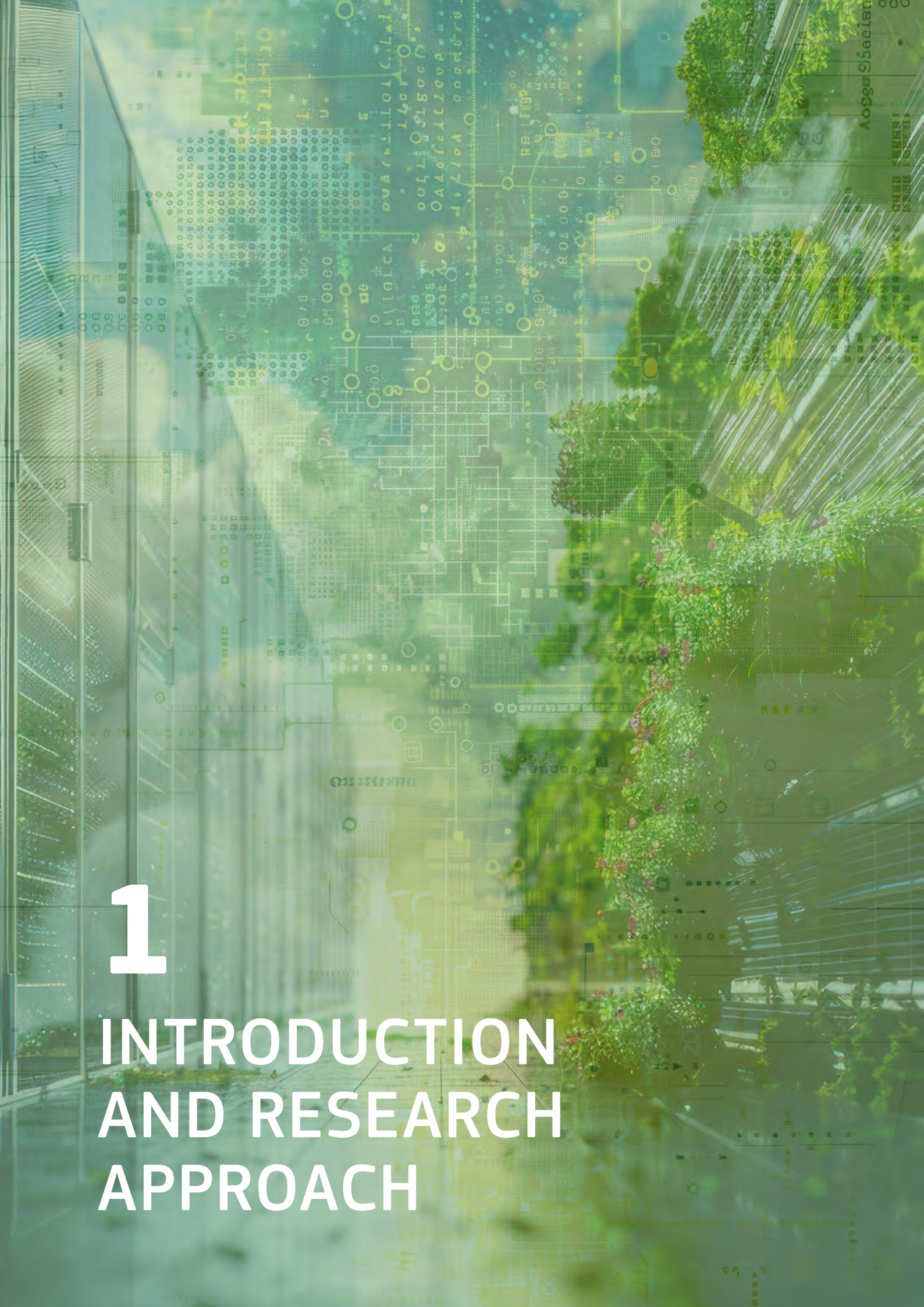
The JRC's experience and expertise in data sharing topics has grown increasingly in the last decades. The JRC has played a **key role in the sharing and exchange of environmental data at the European**

level, being the technical coordinator for the implementation of the INSPIRE Directive since its entry into force in 2007. The present report will inform the revision of this Directive, which will take place under the GreenData4All initiative, planned to be concluded by 2025.

Furthermore, the JRC has published **numerous academic and policy publications on the implementation of the European Strategy for Data**. Among those are the Science for Policy Report on the establishment of common European data spaces, the Report on emerging models of data intermediation in the context of the Data Governance Act, and the Technical Report on consent management tools for data altruism. This work is expected to be expanded in the upcoming years, in alignment with the new priorities of the JRC Work Programme 2025-2027, where both EU innovation and Green Deal ambitions are among the main pillars.

Both the ambitious policy initiatives outlined in the European Strategy for Data – mainly the establishment of common European data spaces – and the objectives of the European Green Deal, necessitate **further scientific research covering technical, organisational, legal, and economic aspects related to data sharing in the EU**. In addition, such scientific evidence is crucial in putting forward a European Data Union Strategy, as envisaged in the 2024-2029 political guidelines of President-elect Ursula von der Leyen.

It would be beneficial for policymakers and the JRC to collaborate in identiying topics for analysis in the areas where common European data spaces can contribute to unlocking and utilising Green Deal data. Specific areas meriting further investigation, as discussed in this report, include the **design of incentives for data sharing and use, legal certainty and simplification, interoperability, trust, and novel business models for scaling data sharing and use**. Experimentation and sandboxing approaches could prove beneficial for such areas of research. Thanks to its longstanding experience, multi-disciplinary competences and research excellence, the JRC is uniquely positioned to continue carrying on such work in the years to come.



1 INTRODUCTION AND RESEARCH APPROACH

1.1 Policy context

The European Union's (EU) pursuit of a fair and trustworthy data economy has been a pivotal aspect of the Commission's policy agenda from 2019 to 2024, and continues to be a central topic in the upcoming 2024-2029 policy agenda, as it has become ever more clear that **'Europe needs a data revolution'** (von der Leyen, 2024). Furthermore, the EU's dedication to advancing a sustainable future is evident in its pursuit of the twin transition, which integrates the green and digital transitions as mutually reinforcing trajectories to enhance Europe's economic competitiveness and resilience. The green transition, guided by the European Green Deal (European Commission, 2019), sets an ambitious path toward climate neutrality by 2050, while the digital transition aims to leverage cutting-edge technologies to stimulate economic growth and societal progress. At the intersection of these transformative agendas lies the imperative for innovative data governance and sharing practices, especially within the ambit of the common European Green Deal data space (henceforth referred to as Green Deal Data Space (GDDS)), which is an integral component of the European Strategy for Data (European Commission, 2020).

In addressing the climate and biodiversity crisis and achieving the targets set by the Green Deal, data serves as a critical resource that impacts decisions on everything from energy efficiency measures to sustainable agriculture, biodiversity conservation, and a circular economy, to name just a few. In an ambitious effort to leverage the potential of digital technologies, the European Strategy for Data envisaged the creation of a single market for data, which would allow information to flow freely within the EU and across sectors for the benefit of businesses, public administrations and societies at large. A pivotal feature of this strategy was the establishment of sector-specific common European data spaces, including a Green Deal data space. In this context, common European data spaces are conceptualised as 'decentralised infrastructures where diverse actors can share, access and use data in a secure, reliable and trustworthy manner, following common governance, organisational, regulatory and technical mechanisms' (Farrell et al. 2023, 10).

However, while the importance of data as a key enabler cannot be overstated, accessing and using them seamlessly and at scale while respecting privacy and security standards, presents several challenges. The sheer volume of data, the complexity of integrating diverse data sets, and the need for robust privacy and security measures are only examples of such **challenges that necessitate novel approaches to data governance and sharing**. Moreover, balancing open access with proprietary interests, and ensuring that data benefits people in an inclusive manner, requires innovative legal and technical frameworks.

This report focuses on emerging strategies to govern and share data within the Common European Green Deal data space. It explores how these strategies can empower stakeholders, including businesses, researchers, and civil society, to contribute to and benefit from a data-driven green transformation. By examining best practices, existing legal and policy instruments, and collaborative models, the report aims to provide actionable insights and recommendations that will enhance the effectiveness of data towards the achievement of Green Deal objectives.

1.1.1 The INSPIRE Directive and the common European Green Deal data space

The policy landscape for governance and sharing of environmentally relevant data within the EU is underpinned by several key legislative and strategic frameworks. On the one hand, sectoral regulations include the INSPIRE Directive (European Union, 2007), which aims to enable access to environmentally relevant spatial information from public sector organisations, and the Directive on Public Access to Environmental Information (European Union, 2003), although the focus of this report is on the former. On the other hand, the European Strategy for Data sets the foundation of Common European Data Spaces from a horizontal perspective (i.e. not limited to environmentally relevant themes, but across domains) and establishes novel regulatory frameworks on data access and governance in general.

The **Directive establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)**, adopted in 2007, was originally conceived to harmonise environmental data across the public sector. It aimed to create a pan-European spatial data infrastructure that would enable the supply and sharing of environmental spatial information among public sector organisations and facilitate public access to spatial information across Europe.

During the last decades, INSPIRE has been instrumental in standardising environmental data and promoting interoperability of spatial data. However, several challenges emerged during its implementation, such as the appearance of parallel implementations in some Member States, and the difficulties of implementing complex technical requirements (European Commission, 2022). In addition, the ever-growing landscape of disruptive technological innovation coupled with the emergence of a new European policy and legal framework for data has resulted in an increased acknowledgment of the need to modernise the Directive. This emerged as a necessary action to meet the evolving demands of digital transformation and to better serve current environmental policy objectives (Kotsev et al., 2021).

Accordingly, the GreenData4All initiative was launched in 2019 to evaluate and review the INSPIRE Directive. The evaluation was completed in 2022 (i.e. see European Commission, 2022) and informed the basis of the review process, which is estimated to be concluded by 2025. The review is expected to address issues related to the limited availability and accessibility of data, the need to simplify technical requirements, and connecting more efficiently the supply and demand of data. The initiative pursues the following specific objectives (European Commission, 2024):

- Make environmental **data available and accessible in a more efficient and future-proof way**.
- Respond to the **needs of environmental policy** development and implementation and enable environmental monitoring and reporting, increase the quality of evidence on the state of the environment and boost the green data economy.
- **Move from a provider-centric to a user-centric approach** where every push for making data available is substantiated by a clear user need creating value.

The revision of the INSPIRE Directive will, therefore, be instrumental to align it with the broader objectives of the GreenData4All initiative, and ensure it supports the ambitions of the European Green Deal data space. As such, it would be important to explore the connections and synergies between the two and take into account the limitations and opportunities that arise from INSPIRE's current focus on the public sector as the main user and provider of data.

FIGURE 1.

Overview of the INSPIRE Directive in the context of the GreenData4All initiative and the GDDS.

	INSPIRE Focus on spatial data sharing between public administration and with the public.	**GreenData4All** Focus on data use and use cases for public authorities, contributing to the Green Deal Data Space.
Interoperability	Covered for environmental spatial data	Environmental data **beyond spatial data**
Reusability	Supply-pushed (data first)	**Demand-driven** (mission first)
Trustworthiness	Focus on spatial data (quality)	Both **in the process and the data**
Accountability	Partially covered	Transparency, responsibility, and compliance with legal and ethical standards.
User-centricity	Too focused on technical side	Designed to meet the needs of **public administrations, business, and citizens**.

Green Deal Data Space

Source: own elaboration.

The **common European Green Deal Data Space (GDDS)** builds upon the European Strategy for Data (European Commission, 2020). The Data Governance Act (European Union, 2022), the Data Act (European Union, 2023) and the Implementing Act on High-Value Datasets (European Commission, 2023) under the Open Data Directive (European Union, 2019) provide the legal framework. In addition, the GDDS is underpinned by sectoral legislation, mainly the previously mentioned INSPIRE (European Union, 2007) and the Public Access to Environmental Information Directive (European Union, 2003).

The **Data Governance Act** aims to increase trust in data sharing and bolster the use of data for societal and economic purposes. The Data Governance Act establishes a regulatory environment that encourages the reuse of public sector data and introduces mechanisms for data altruism, where individuals and companies can share data voluntarily for the common good. It also proposes the concept of data intermediaries to facilitate data sharing while ensuring compliance with EU values and regulations. Complementing the Data Governance Act is the **Data Act**, which seeks to address issues related to data access and data portability with a focus on data generated from machines and the Internet of Things (IoT), with the objective of unlocking the value of data across different

sectors. The Data Act aims to establish clear rules on data sharing and usage, promoting fairness and fostering innovation by levelling the playing field for businesses, especially small and medium-sized enterprises (SMEs) and start-ups, and by granting public-sector bodies and EU citizens more control over their data. Furthermore, it also complements the GDPR (EU General Data Protection Regulation, European Union 2016) through the extension of access regimes for non-personal data to those on personal data under the GDPR. Finally, the **High-Value Datasets Implementing Act** (European Commission, 2023) further builds upon these efforts by identifying specific public-sector datasets that hold considerable potential for generating socioeconomic benefits from their reuse. These datasets are to be made available for free and under open licenses across the EU, thus enhancing their value and stimulating innovation and growth in the data economy.

The mission of the GDDS is to leverage the regulatory provisions – such as those mentioned above – plus state-of-the-art technology in order to boost an economy of 'green' data that will in turn contribute to achieving the objectives of the Green Deal, especially in the areas of a circular economy, biodiversity, zero pollution and climate change/adaptation.

However, fully leveraging the opportunities presented from the green data economy requires a transition from fragmented approaches to the inclusive participation of a diverse array of stakeholders (Thabit and Mora, 2023) – from SMEs, to citizens, nonprofit entities, data intermediaries, and data altruism organisations. It also requires a stronger connection with the effective utilisation of data, i.e. data-driven business models that create value.

2.2 Towards an impact-driven view: The GDDS as an emerging system

Kotsev et al. (2021) argue for the need for spatial data infrastructures, including the one stemming from the INSPIRE Directive, to *evolve from complex and highly specialised geospatial data frameworks, where legal obligations are enforced by strict technical specifications, to flexible, open, agile, and self-sustainable data ecosystems* (p.13).

The ultimate goal is therefore to foster the emergence and development of such data innovation ecosystems. During the last decades, academic research has strongly advocated for the need to adopt a systemic view of innovation, which focuses on understanding the complex interactions within entire systems rather than isolating individual components (Midgley and Lindhilt, 2021). Within this view, a data innovation ecosystem generally involves a network of elements and entities that interact and are integrated in each area within several infrastructures that are involved in the creation, transference and use of data (Chae, 2019). In the context of INSPIRE and the GDDS, these network 'elements' are represented by actors (e.g. data intermediaries, data altruism organisations, citizens, businesses, public authorities); they exist within legal, cultural, social, and governance mechanisms that underpin the architecture of the system defining the rules and possibilities for the actors to perform their roles.

Another characteristic of data spaces is their high level of complexity both in terms of data and interoperability (Solmaz et al., 2022), as well as because of the amount and diversity of entities involved (Dashmukh et al., 2023). System thinking is a well-demonstrated and

suitable way to tackle complex systems (Monat and Gannon, 2015; Hossain et al., 2020).

The implications of adopting a systemic approach in researching the future of INSPIRE in the context of data spaces are primarily:

- **The main focus is on value and impact from data use.** As opposed to assuming a supply-centric view, systems thinking leads to considering the whole as a 'data economy' – with supply and demand – where data is the 'product' used to create value (economic, social, etc.) exactly as products create value in the 'real' economy.
- **This means to practically complement the so far dominant 'supply-pushed' view of data ecosystems, with a 'demand-driven' perspective,** consistent with data use and value creation being the ultimate goal of the GDDS. The 'demand-driven' perspective is expected to encourage more and more actors to perform relevant roles within the Green Deal data economy.
- As a consequence of the previous point, the focus is on the **sustainability of data sharing and use over time**. Embedding a demand-driven perspective through system thinking entails emphasising the need for data to be shared, under certain conditions, with continuity over time. The phenomena of interest and the objectives of the Green Deal, besides often requiring a longitudinal approach, typically benefit (near) real-time data. It is therefore crucial to foster a system of data generation – sharing – use dynamics, whereby data flows are sustainably established enabling the emergence of self-organised systems of data-enabled impact on the Green Deal.
- **Focus on clear objectives to be achieved in a specific domain.** As argued in more detail below, all processes, entities, technologies, and regulatory provisions embedded in the GDDS should be seen as a 'means to an end', i.e. instrumental to removing barriers that prevent the 'green' data-economy from performing optimally and achieving the Green Deal objectives. These barriers include a lack of systems to collect the data needed, lack of data interoperability, fragmentation, lack of data portability, data lying behind intellectual property and confidentiality 'walls', and many others.
- Shift from looking at the ecosystem as the sum of discrete and individual actors and roles, to **tackling the ecosystem as a whole**, thereby focusing on interrelationships and interconnections between actors and roles.
- Acknowledging that the emergence of the GDDS is, to some extent, **the result of self-organising processes** where actors perform their roles, influenced by their own (often not aligned with others) motivations, interests, and capacities. This self-organisation process happens in each legal and technological context. The extent to which the latter should be flexible (i.e. allow higher or lower levels of self-organisation) is a key question for the development of the GDDS.

This view is very much consistent with the key objectives of existing data ecosystems (Curry, 2020; Otto et al., 2022), which focus on clear objectives to be achieved in a specific domain, as opposed to emphasising data attributes and characteristics. As an example, the European Mobility Data Space (Montero-Pascual et al., 2023) has been established to pursue three overarching goals: reduce

greenhouse gas emissions from transport, digitalise mobility services and processes, and foster resiliency of mobility (Li and Quinn, 2024). The vision is therefore to achieve a system whereby data interoperability, justice, and sovereignty will be achieved through (data) governance mechanisms and a consistent reference architecture.

Furthermore, adopting a systems thinking approach means addressing two prominent gaps in data ecosystems. First, producing, sharing and using data takes significant effort due to the variety of native formats in which data are produced and stored by the various stakeholders and associated licenses, the different standards adopted, a lack of alignment of actors' objectives and attitudes, and the complex, often unclear value network within the ecosystem. The result is that, in most cases, data ecosystems appear to be scattered, fragmented and poorly understood beyond individual data flows and use applications, i.e. lacking a systemic approach. The latter leads to focusing on interrelationships and interconnections between actors and roles, beyond their identification and mapping, as argued above.

Second, a systems thinking approach addresses the limitations of focusing on a supply-driven approach, which has been dominant to-date, especially within the open data developments and initiatives. In this paradigm, effort and resources have been invested primarily in publishing data according to certain characteristics, such as the widely adopted five-stars model developed by Tim Berners-Lee (2006). Taking a systems thinking approach entails defining the desired output as data use (to lead to the outcome of positively impacting the Green Deal), as opposed to a focus solely on sharing and publication of data. A systems thinking approach aligns with theories on Earth System Governance (Biermann 2007). When applied to considerations regarding the GDDS, this ensures that governance of Green Data considers different aspects including the role that institutions such as governments, civil society and businesses play in determining policy regarding data use in a data economy. It also includes the accountability and legitimacy of these institutional structures in data economies and how they adapt to environmental changes as well as policy changes. Further it also addresses the ways in which resources are distributed amongst actors to ensure fair outcomes in relation to data use taking the power relations between different actors into consideration. To summarise, Figure 2 provides a general representation of the GDDS system, which exemplifies how a systemic approach leads to shift from considering a process of data creation – data sharing – data use, to considering the GDDS as a system characterised by positive loop whereby:

- Value from data use (i.e. impact on the Green Deal), together with institutional, social, environmental, economic and legal pressures, drives the development of vision, commitment and incentives to data creation and sharing. Legal pressures are particularly relevant for driving data supply. Existing laws and directives embed a wide variety of pressures (i.e. drivers for data supply) such as the open-data movement, data altruism, open environmental data, public investments (e.g. in the case of the INSPIRE geoportals), open science principles, among several others[1].
- Legal pressures (driving data supply) are reinforced and complemented by those mechanisms driving the demand for

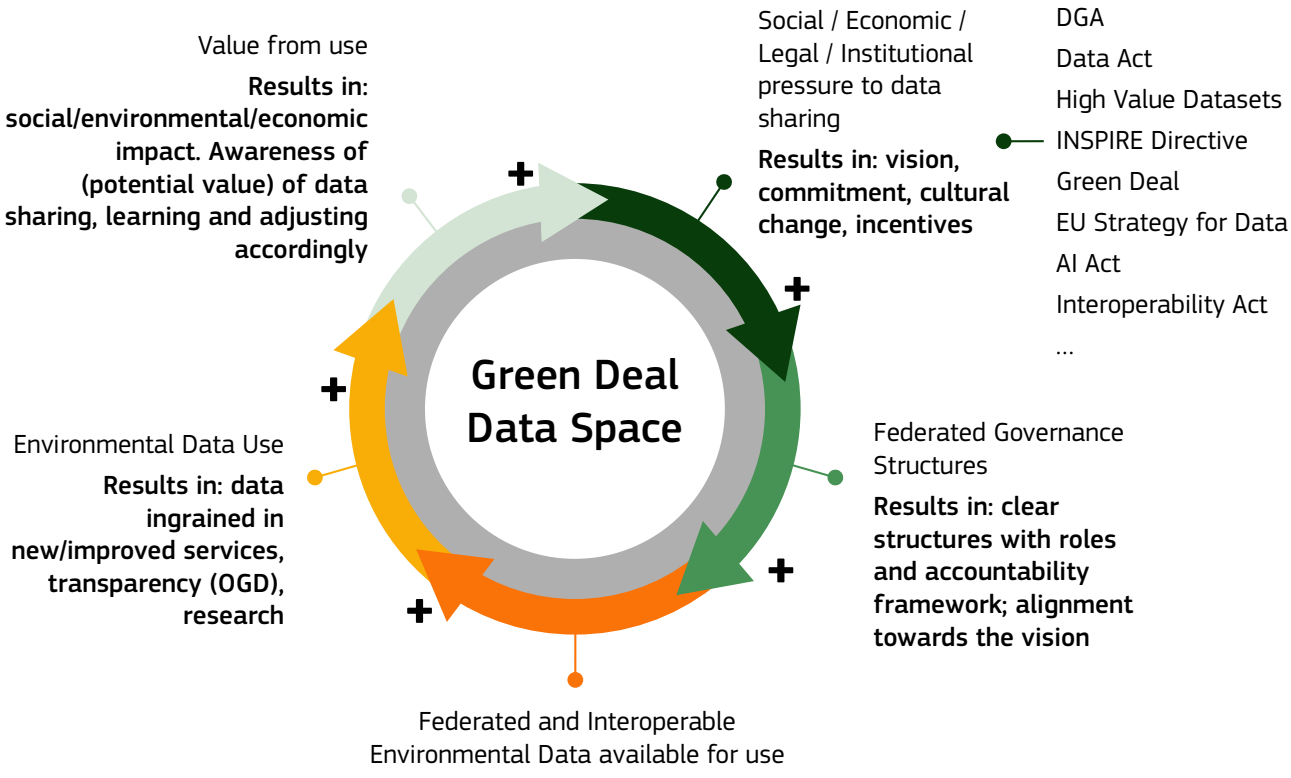
1. See Chapter 2 for a more extensive analysis of incentives and disincentives associated with the existing relevant laws and directives.

data. Examples are seemingly countless as these depend on several societal, cultural, and environmental factors. They may be the strongly advocated need to improve our environmental awareness (e.g. for forests, soil and other climate related trends or events), demand for data for business/operational intelligence, data as a proof of compliance with due diligence regulations, data for transparency purposes to lure customers and consumers (marketing), data for market transparency to foster circular economy principles (e.g. in the textile waste market), among many others.

- These pressures, especially those provided by EU regulations and directives, positively contribute to the creation of appropriate governance mechanisms. These are established to promote clear and transparent accountability frameworks while defining policies that foster alignment towards the vision (i.e. the Green Deal itself).
- These emerging and established structures self-organise and create a data architecture and infrastructure where data is shared through appropriate data governance schemes and access points, as well as being consistent with certain standards, attributes, and characteristics.
- These foster data use, i.e. the creation of impact of the Green Deal through embedding data in new or improved services, for research and/or for transparency thus creating additional pressures driving this positive loop.

FIGURE 2

A systemic approach to the GDDS: connecting supply and demand.



Source: own elaboration.

It should be noted that while Figure 2 is meant to be neither granular nor prescriptive, it provides a consistent view of the shift promoted by adopting systems thinking in this study, i.e. from seeing the GDDS as an ensemble of linear 'input-output-outcome' processes, to a complex ecosystem characterised by causal loops and dynamics. From this perspective, important aspects such as achieving interoperability, creating incentives, researching effectively, and just governance mechanisms for data collection and stewardship, among others, are therefore instrumental for achieving Green Deal objectives, rather than being ultimate goals in themselves.

It should also be noted that this systemic view resonates with the ultimate goal of fostering an effective data economy across the EU and beyond. The GDDS system indeed follow the same principles of traditional economies, primarily built upon supply and demand (i.e. in this case 'data supply') mechanisms. The resulting data economy enabled by the GDDS, as with any other economy, will create (data) products and (data-enabled) services; it will establish (data) supply chains across actors and through intermediaries; and trust and relational governance mechanisms (e.g. contracts and licenses) will be crucial. In other words, the GDDS will deploy all the necessary technology/regulatory provisions enabling continuous mechanisms reinforcing supply and demand (and vice versa) thus creating sustainable data sharing and use, i.e. an effective data economy to support the implementation of the Green Deal.

1.3 The need for a fair and equitable Green Deal data space: A data justice perspective

The European Green Deal focuses on the decoupling of economic growth from resource use and the resulting environmental impact. However, it emphasises that the process of developing a competitive economy that is also resource efficient must be carried out in a manner that leaves no person or place behind (European Commission 2020a).

In its approach to developing this outlook, the Green Deal explicitly states that it is 'more than setting targets' (European Commission 2020a). The agenda argues for **'putting people at the core of the transition'** by ensuring that such transitions prioritise justice and fairness and supports those who are most vulnerable during such a transition, as well as those affected by climate change (European Commission 2020a). With the articulation of the European Strategy for Data (hereinafter the 'Strategy'), which led to the development of the Common European Data Spaces, significance is placed on developing both data infrastructure as well as governance that results in greater data pooling, access, and sharing and ultimately use of data. The strategy is designed to allow for stakeholders to drive the evolution of the spaces.

The Data Spaces are designed to be 'open' through encouraging participation, 'secure' with a focus on privacy to ensure that data can be shared, and 'practical' to ensure access and use of data through rules that are fair and enable trustworthy data governance. Further, the Strategy not only seeks to empower and encourage data holders to make their data available for reuse, but also seeks to ensure

that EU rules and values regarding personal data and consumer protection are adhered to (European Commission, 2020). In keeping with these motivations, the GDDS is designed to ensure that data can **facilitate greater environmental transparency, improve decision-making by public authorities, and develop informed citizens** (European Commission 2024).

To align goals of justice, fairness, empowerment and access that are present in the Green Deal and the Strategy, and ensure that they can be fulfilled, we propose to draw from a data justice perspective to frame an approach to GDDS. Data justice is an approach that places a priority on concerns of social justice as they manifest in debates around ethics, human rights, or forms of activism. The influence of data extends beyond its collection and utilisation, shaping societal norms through the justifications it fosters. These justifications arise from how we perceive data, e.g. whether as a valuable resource for extraction or a tradable commodity or for instance in terms of social relations (Dencik & Sanchez Mondero 2022, Viljoen 2021).

Through a data justice lens, it is possible to explore overlooked aspects of the development of data spaces, such as representation, inclusion, accountability, and agency (Taylor 2017). It is critical to challenge the ways in which structural injustices are being replicated through datafication processes, and consequently requires looking at data through an assessment of the socio-political and material contexts rather than as a technocratic exercise (Dencik, Jansen & Metcalfe 2018). This entails examining the context in which data are being used, rather than purely examining data in terms of technical aspects without the wider socio-political considerations.

In the context of the Green Deal and the generation of green data, an important framing that is relevant from a data justice standpoint (Taylor, 2017), is the following:

1. How is data bringing attention to the concerns of different stakeholders, and how are all of their concerns being represented?
2. In what ways does the involvement with data ensure that people are empowered to share in its benefits, and have the choice and autonomy to determine how it is used?
3. How can people challenge and prevent biases and discrimination in data-driven processes including governance or policymaking based on data?

The concept of fairness from a data justice perspective in the context of green data can also be understood across three axes. As Heeks and Renken (2016) argue, this can be: (i) instrumental, where attention is paid to fair use of data and less, for instance, on who produces and owns the data; (ii) procedural, which examines the fair handling of data, where one studies how data are captured, inputted, processed, and even used; and (iii) in terms of distribution and distributive effects, which focuses more on data asymmetries, looking at who controls the data and who has rights to it (Heeks and Renken 2016).

With governance being a key consideration of the Strategy, ensuring that these are justice-oriented requires evaluating firstly to what extent does the governance structure benefit the public, and public infrastructures. Secondly, does it account for the rights of people in inclusive ways, considering their differentiated contexts and lived experiences. Thirdly, does it have modes for redress when people have grievances, and finally to what

extent governance mechanisms play a role in influencing, contexts and markets outside the EU (Lopez Solano, Martin, de Souza and Taylor 2022).

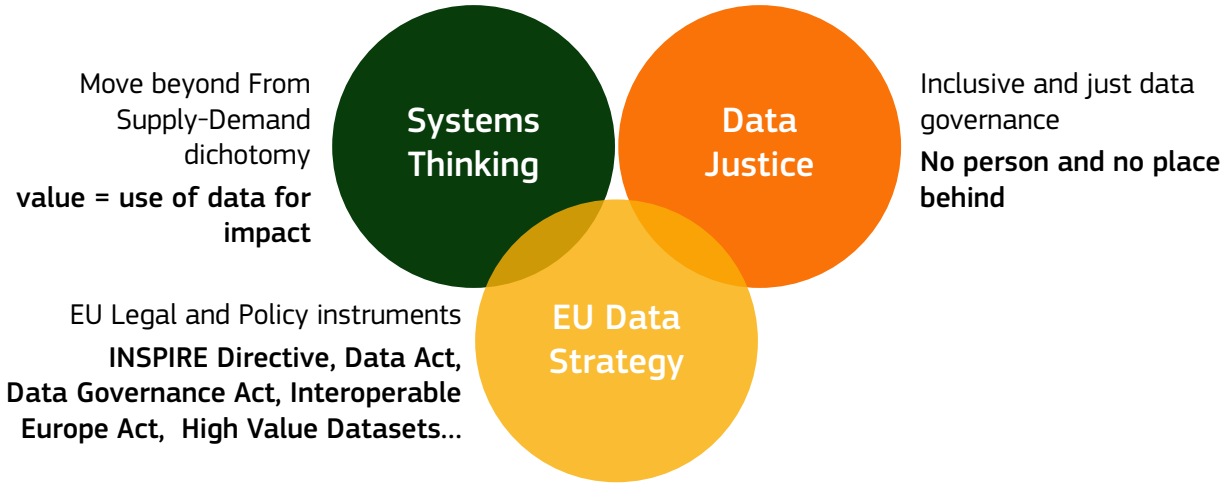
Through adopting a data justice standpoint, this report is concerned with positioning the strategy on Green Deal data spaces as one that studies the political economy behind which data are generated and owned, the implications of data distribution, and the ways in which what is considered as legitimate data is imagined and determined (Vera et al 2019).

1.4 Approach and structure of the document

The current study is the result of a collaborative effort among an interdisciplinary group of experts from various universities and research organisations, encompassing fields such as computer science, engineering, law, and sociology. Aiming at exploring novel perspectives on emerging approaches for data governance and sharing, within the context of the revision of the INSPIRE Directive and the implementation of the common European Green Deal data space (GDDS), the research is structured around the three main pillars outlined in preceding sections. Firstly, the opportunities and challenges presented by the ambitious legal and policy frameworks stemming from the European Strategy for Data (European Commission, 2020). Secondly, the application of systems thinking to underscore the significance of aligning supply with demand and concentrating on impactful outcomes. And thirdly, the integration of a data justice lens that emphasises the need to ensure that no person and no place is left behind in the green and digital transition.

FIGURE 3

A systemic approach to the GDDS: connecting supply and demand.



Source: own elaboration.

The subsequent three chapters of this report focus on the opportunities and challenges brought by the horizontal provisions that underpin data spaces, specifically the Data Governance Act (DGA) and the Data Act, from various perspectives. **Chapter 2**, 'Governance structure for environmentally relevant data sharing: An analysis of regulatory

incentives and disincentives in the European Green Deal data space', written by Marisa Ponti, investigates how the DGA and Data Act, together with the sectoral legislation underpinning the GDDS (i.e. INSPIRE and Public Sector Environmental Information Directives), could incentivise data sharing across various actors in the data space. **Chapter 3**, 'Environmental data intermediation services: Mapping the Green Deal data spaces ecosystem', by Manuel Portela Charnejovsky, explores different roles and modalities of data intermediation services, which are one of the central actors conceptualised in the DGA. With a focus on value creation, it provides an overview of key intermediary actors of environmental and geospatial data, and unveils interconnections and similarities between the GDDS and other common European data spaces. The report continues with **Chapter 4**, 'Data altruism and the Green Deal data space: Mapping the ecosystem', where Paola Pierri examines the approaches and models of data altruism in the context of the DGA and provides insights on the key actors and actions that are critical for this model to be successfully implemented.

The study follows with an analysis on the role of non-traditional players in the governance, sharing, and utilisation of geospatial and environmental data. In **Chapter 5**, 'Business data sharing and the Green Deal data space', Angela Daly examines the participation of, and data contribution from businesses, from a public good perspective. She assesses opportunities and challenges for B2G, B2B and B2C data sharing, leveraging corporate social responsibility and environmental data sharing, alongside challenges coming from various legal regimes. **Chapter 6**, 'Citizen-generated data within the Green Deal data space', written by Stefania Milan, offers an actionable definition of citizen-generated data (CGD) positioning citizens as knowledge makers and data users in the GDDS, while reviewing best practices to foster and support the generation and utilisation of CGD. Additionally, **Chapter 7**, 'Implementing CARE principles in data sharing with Indigenous peoples and local communities', developed by Riikka Kaukonen, provides a deeper dive in the topic of CGD governance and data sharing, and emphasises the importance of CARE principles[2] to ensure fair and sovereign approaches.

The report closes with **Chapter 8** on key conclusions and policy recommendations from a Systemic data justice perspective. Developed by Giovanni Maccani, Siddharth Peter de Souza, and Sara Thabit González, it provides strategic advice and actionable points for the revision of INSPIRE and the implementation of the Green Deal data space.

2. Collective Benefit, the Authority to Control, Responsibility and Ethics.

2 GOVERNANCE STRUCTURE FOR ENVIRONMENTALLY RELEVANT DATA SHARING

AN ANALYSIS OF REGULATORY INCENTIVES AND DISINCENTIVES IN THE EUROPEAN GREEN DEAL DATA SPACE

2.1 Introduction

2.1.1 Policy context

In the Communication 'A European strategy for data' (European Commission, 2020), the Commission aims at supporting the creation of a Common European data space, to use the potential of data to support the Green Deal, a wide-ranging set of policy initiatives aimed at transforming the European economy to be more sustainable, reducing greenhouse gas emissions, and promoting economic growth that is both environmentally friendly and socially inclusive (European Commission, 2019). Recently launched in the context of the European Data Strategy are also the Data Governance Act (European Union, 2022) and the Data Act (European Union, 2023) – the two horizontal layers of data regulation that aim to facilitate data sharing and reuse across various sectors.

To facilitate the implementation of the Green Deal strategies and the achievement of Sustainable Development Goals (SDGs), the European Green Deal data space (hereinafter GDDS) is expected to evolve into a shared platform for data, algorithms, and applications used by businesses, consumers, government institutions and citizens alike, fostering connectivity between public and private data environments. The intention is to provide an interoperable and trusted space to connect currently fragmented and dispersed data from the private and public sectors by, among other objectives, drafting a set of rules of 'legislative, administrative and contractual nature' that determine the rights of access to and use of data (European Commission, 2021). The GDDS is tailored to enhance the accessibility and use of environmental observational data and, similarly to other Common European data spaces, is envisioned to foster 'an ecosystem (of companies, civil society and individuals) creating new products and services based on more accessible data' (European Commission, 2020, p.5).

As highlighted in their working document on Common Data Spaces, the Commission (2022b) specifies that data spaces encompass more than just the technological deployment of data sharing tools and services, and the integration of energy-efficient and trustworthy cloud capacities and related services. They also underscore the importance of data governance structures, compliant with relevant EU data legislation (e.g. the Data Act and Data Governance Act), and which transparently and fairly delineate the rights of access to and processing of data. This understanding underscores the necessity of integrating customised legislation and governance mechanisms to guarantee data accessibility. By implementing tailored legislation, robust governance mechanisms, and investing in necessary infrastructure and competencies, the EU aims to create an environment that encourages businesses and the public sector to store and process data within its jurisdiction. In the Strategy for Data, the Commission (2020) points to the importance of developing **appropriate incentives to create a data sharing culture**. This highlights the role that regulatory incentives created by EU laws play within a governance model of the GDDS, thus ensuring the realisation of its full potential. While it is evident that these laws influence behaviour, the central question revolves around **optimising their effectiveness**. This entails ensuring that these interventions successfully advance environmental, social, and economic objectives while remaining economically viable for society.

A key premise is that a governance of the GDDS should expand **beyond the technologically focused concept of a federated platform of data and services**, which primarily facilitates data sharing and reuse among various public and private platforms, towards **a more ecosystemic approach centred around federated governance of the whole GDDS**. This shift prioritises not only technological interoperability standards but also other crucial governance elements such as multi-stakeholder participation, clear distribution of responsibilities, effective regulatory incentive mechanisms, and consideration of the social and economic implications of data sharing and reuse for the stakeholders. This ecological form of federated governance can help navigate better the complexities of the GDDS governance and foster a **more inclusive and sustainable data sharing ecosystem**.

2.1.2 Purpose and objectives of this chapter

This chapter addresses the following questions:

- ***What regulatory incentives do relevant EU data directives and regulations create for accessing and sharing environmentally relevant data?***
- ***What regulatory disincentives arising from these EU data directives and regulations can discourage actions aimed at complying with the requirements for accessing and sharing environmentally relevant data?***

The relationship between governance systems and incentives significantly shapes the dynamics of accessing and sharing environmentally relevant data in compliance with the EU data legislation. Environmentally relevant data encompasses not only data directly related to the state of the environment, such as air and water, but also data likely to impact the state of the environment, e.g. road traffic and use of fertilisation for crop management.

By examining regulatory incentives and the potential disincentives for accessing and sharing environmentally relevant data, the analysis aims to contribute to understanding how the selected EU data legislation can induce changes in the behaviours of individuals and private and public organisations to produce the environmental, social, and economic benefits prescribed by these laws. In turn, this analysis is poised to inform the purpose of integrating incentives and other non-regulatory interventions into the governance structure of the GDDS.

2.2 Scope and method

EU data legislation does not differentiate between data relevant to the environment and data unrelated to the environment. It has been noted that, despite several existing regulations governing data access, particularly in environmental domains, a clearly defined, standardised legal framework for environmentally relevant data is lacking (Finck & Mueller, 2023).

This chapter does not aim to conduct a comprehensive review of all EU data legislation. It rather focuses on four cross-sectorial directives and regulations (Table 1) that are particularly relevant to the GDDS:

TABLE 1.

EU Directives and Regulations included in this study.

Legislation	Purpose
Data Governance Act Regulation (European Union, 2022)	It aims at increasing trust in voluntary data sharing for both economic and public interest purposes, with a focus on public sector data, data intermediation services, and data altruism organisations.
Data Act Regulation (European Union, 2023)	It aims at enabling a fair distribution of the value of data and making more data available to companies, citizens and public administrations.
Freedom of Access to Information Directive (European Union, 2003)	It sets the right of access to environmental information held by or for public authorities and ensures that such information is made available and disseminated to the public.
INSPIRE Directive (European Union, 2007)	It aims to create a European Union spatial data infrastructure (SDI) for EU environmental policies and policies or activities that may impact the environment.

Source: own elaboration.

Reasons for inclusion: the Data Governance Act (DGA) and the Data Act have been presented to take environmental concerns into account and therefore seem most relevant for access to both public and private sector data for ecological purposes. For example, in the Data Act, under 'Reasons and Objectives for the Proposal, 1. Context of the proposal' in the Explanatory Memorandum of the Data Act (European Union, 2023), data is described as 'an essential resource to secure the green and digital transitions', while under 'Consistency with existing policy provisions in the policy area', the initiative is presented as aiming to enable 'EU public authorities, businesses and citizens to support the transition to a greener and carbon-neutral economy and reducing administrative burden'. In addition, the Freedom of Access to Information Directive and the INSPIRE Directive can be evaluated and possibly reviewed from an environmental angle. The Freedom of Access to Information Directive aims to improve environmental protection by increasing public access to and the potential dissemination of environment-related information held by the competent authorities of the Member States or by EU institutions. The INSPIRE Directive aims to create a better basis for environmental policymaking and establish a common spatial data infrastructure, which is important to address data sharing obstacles from the past and data needs for the future.

Reason for exclusion: excluded are sectorial regulations (e.g. mobility, health, etc.) that cover environmentally relevant data, or even cross-sectorial regulations, such as the Open Data Directive that covers only the public sector. For this analysis, we gave priority to general data acts as they provide a foundational framework that applies to a broad range of data across various sectors. These horizontal layers can be further complemented with domain or sector-specific data-related obligations through other 'data spaces'.

2.2.1 Analysis

The analysis was conducted with a high level of granularity. This decision was driven by the need for a general understanding of the selected legislation, while optimising the use of time and resources. Analysing

content at a low granular level poses the risk of getting lost in excessive detail, potentially obscuring overarching patterns and themes. By opting for a higher level, the aim is to efficiently identify some key themes across the four acts. This approach allows sifting through data more efficiently, distilling it into manageable insights that provide a comprehensive overview. Furthermore, analysing at a higher level facilitates easier comparison and synthesis of findings, aiding in the formulation of overarching conclusions and recommendations. The analysis was conducted using systematic content analysis of the directives and regulations. Systematic content analysis has been used in legal research (Hall & Wright, 2008). This method does not replace conventional legal analysis, as systematic content analysis ignores aspects of legal interpretation that look at, for example, nuances related to infrequent or highly complex factual and procedural patterns (Hall & Wright, 2008). Systematic content analysis requires the following steps: formulating the research questions to be answered, selecting the sample to be analysed, defining the codes to be applied, condensing raw data into concepts based on sound interpretation, determining trustworthiness, and analysing the results of the coding process (Hsieh & Shannon, 2005). Interpretation of text was conducted to avoid compromising the original meaning expressed in the legislation and ensure reliable inferences.

The data analysed for this study included the following chapters:

- **Data Governance Act:**
  - Chapter II on reuse of certain categories of protected data held by public sector bodies (Arts. 3-9);
  - Chapter III on requirements applicable to data intermediation services (Arts. 10-15);
  - Chapter IV on data altruism (Arts. 16-25).
- **Data Act:**
  - Chapter II on business-to-consumer (B2C) and business-to-business (B2B) data sharing (Arts. 3-7);
  - Chapter III on obligations for data holders to make data available pursuant to union law, which include compensation and technical protection measures (Arts. 8-11);
  - Chapter V on making data available to public sector bodies on the basis of an exceptional need (Arts. 14-22);
  - Chapter VI on switching between data processing services (Arts. 23-25).
- **INSPIRE Directive** Chapter V on data sharing.
- **Freedom of Access to Information Directive:** Chapter on access to environmental information upon request (Arts. 2-7).

Given time limitations, the recitals providing the introductory part of this legislation have not been included in content analysis, as they are not legally binding. The study is a small and time-limited preliminary assessment, aimed at identifying potentially relevant incentives and is subject to further refinement and expansion. Thus, this assessment helps

explore and gather preliminary data to inform the development of a larger study. While efforts have been made to provide valuable insights, a more thorough examination is recommended for a more comprehensive and in-depth understanding.

2.3 Background: environmentally relevant data, regulatory incentives, and federated governance of the data space: connecting the dots

For decades, economic literature has explored the notion of incentives across diverse branches of the field, ranging from macroeconomics to financial and behavioural economics. In economics, an incentive is a potential gain, while in public policy it can be seen as a method for changing behaviours. Incentives play a critical role in the decision-making processes of individuals, firms, and governments, prompting extensive examination across diverse work and social contexts. They can be positive or negative and can take various forms such as financial rewards, recognition, social approval, or avoidance of punishment. According to Ostrom (2005), incentives extend beyond simple financial rewards and penalties. They encompass the positive and negative changes in outcomes that individuals anticipate from their actions within a specific set of rules, influenced by physical and social contexts. Additionally, they can be direct, such as reducing the cost of a service, or indirect, such as unemployment benefits incentivising people not to actively look for a job. Actors act and interact based on these incentives, with their activities, in aggregate, generating patterns of outcomes.

In economics, the study of incentives is crucial for understanding how individuals respond to changes in prices, policies, regulations, or other economic stimuli. For instance, if a law provides subsidies to make health insurance more affordable, people are incentivised to purchase health insurance coverage. Analysing incentives within formal rules and public policies is also important, given the role played by rules and policies in regulating interactions among individuals. Laws and formal rules have been defined 'obligations backed by incentives' (Galbiati & Vertova, 2014). Understanding the interplay between incentives and obligations, which constitute the essence of norms, is considered essential in shaping the development of public policies (ibid.). Laws establish systems of rewards and penalties for actors. Regulatory incentives are policy instruments and integral components of governance systems as they can be used to influence, directly or indirectly, behaviours and promote desired outcomes. The results can be dysfunctional, but laws cannot avoid creating incentives of some type.

In the governance of the GDDS, incentives for data sharing and reuse are of great importance because environmentally relevant data serve not **only as an economic good but also to fulfil the public interest**. From this twofold perspective, environmentally relevant data may be regarded in two ways (Purtova & Van Maanen, 2024):

- as a 'club good', which is excludable to some extent and non-rivalrous; **or**
- as a 'knowledge common' – a collective resource accessible to and shared by a wide array of stakeholders, including citizens who can share their data for

altruistic purposes. Data as a knowledge common should be held through a partnership, a not-for-profit, or other entity, for the benefit of all stakeholders, but not owned privately for commercial gain.

Looking at environmentally relevant data and governance of the GDDS through the prism of either data as knowledge commons or as a club good, one needs to construct a form of governance scheme that leverages regulatory incentives. This approach underscores the importance of aligning regulatory mechanisms within a federated governance structure of the GDDS. Such a structure not only emphasises the technical infrastructure for federating data and services, aimed at enhancing cross-border interoperability (European Commission, 2022a, p.7) and ensuring the foundational elements of data spaces (European Commission, 2022b), but also recognises the important role of regulatory incentives in driving desired behaviours and outcomes. Federated governance decentralises authority, distributing decision-making among multiple autonomous centres. In the context of GDDS governance, this decentralisation aims to strike a delicate balance between centralised coordination and local autonomy. Regulatory incentives play a central role in incentivising cooperation and adherence to governance principles across the federation. By strategically leveraging these incentives, the governance framework seeks to foster efficient collaboration while accommodating the diverse and specific needs of stakeholders within the GDDS ecosystem.

2.4 Findings

2.4.1 Overview of regulatory incentives and disincentives

This section summarises the main findings from the qualitative analysis of the selected directives and regulations. Rather than providing exhaustive detail, the focus is on highlighting key insights. The analysis has been influenced by institutional analysis[3] but has not strictly adhered to all its principles or methodologies. A crucial component of institutional analysis entails identifying the incentives faced by individuals and organisations and understanding the origin of such incentives. In this study, incentives originate from the examined data laws. However, identifying the full range of incentives is challenging, and the analysis is by no means exhaustive. Particularly, the focus is on positive and negative incentives faced by stakeholders – especially data holders – due to the provisions governing data sharing and reuse. The identification of incentives is influenced by the nature and attributes of the rules, including their clarity, enforceability, and flexibility. Table 2 displays the main negative and positive incentives found in the data.

3. Institutional Analysis and Design (IAD) has been developed since the 1950s by scholars and practitioners associated with the Workshop in Political Theory and Policy Analysis, Indiana University, Bloomington, IN, USA. IAD is used to analyse a multitude of problems including institutional issues in the governance and management of renewable resources such as forests, fisheries, groundwater basins, watersheds, and pastures.

TABLE 2.

Regulatory incentives and disincentives for accessing and sharing data in the GDDS.

Regulatory incentives	Definition	Type
Neutrality requirement and structural separation	Legal requirements to ensure neutrality and legal independence in intermediation services, avoid conflict of interests for providers of intermediation services, and to ensure legal independence of data altruism organisations from any entity that operates on a for-profit basis.	Direct
Financial compensation	Monetary benefit offered to encourage behaviour or actions which otherwise may not take place.	Direct
Portability requirement	Obligation of providers of connected products and data processing services to enable customers' adoption, development, or use of specific technologies.	Direct
Public security guarantees in data transfer agreements	Protection from disclosure and unauthorised access due to concerns of societal security. E.g. releasing certain sensitive geodata can impinge on national security.	Indirect
Commercial guarantees in data transfer agreements	Protection of sensitive information to maintain a competitive advantage or safeguard proprietary data.	Indirect
Regulatory disincentives	Definition	Type
Legal uncertainty	Unclear formulations in laws that create challenges for various stakeholders. E.g. ambiguous legal language that is open to interpretation and may lead to confusion, disputes, and difficulties in compliance.	Indirect
Manufacturer technical control	Control of the technical design of connected products and related services, exercised by the manufacturers of those products.	Indirect
Transaction and compliance cost	Costs incurred when buying or selling a good or service, and preparing and enforcing a contract, including legal fees, negotiation and monitoring costs, commissions, and other costs associated with conducting business transactions and complying with regulation.	Direct

Source: own elaboration.

Tables 3 and 4 indicate, respectively, the number of occurrences of regulatory incentives and disincentives found in the selected legislation. Table 3 shows three incentives being present across the legislation, indicating their cross-policy relevance. Specifically, both public security guarantees and commercial guarantees in data transfer agreements underscore their fundamental roles in regulating the exchange of data, where such guarantees are necessary to protect confidentiality provided for by national or European law. This recognition underlines the need for comprehensive and adaptable regulatory frameworks that balance security, privacy, and economic considerations in governing environmentally relevant data sharing. In addition, the presence of financial compensation across all the selected legislation underscores its relevance and potential effectiveness as a mechanism for incentivising the reuse of environmentally relevant data sharing, either to support objectives of public interest or increase public access to information on the environment.

By contrast, Table 4 illustrates the cross-policy relevance of the three disincentives

by showcasing their occurrences across two regulations. Disincentives such as legal uncertainty, manufacturer technical control, and transaction and compliance costs are found in the DGA and the Data Act, suggesting their significance in influencing behaviours and decisions related to environmentally relevant data sharing in these two laws.

TABLE 3.

Comparison of regulatory incentives based on frequency.

Legislation	Neutrality requirement and structural separation	Financial compensation	Portability requirement	Public security guarantees in data transfer agreements	Commercial guarantees in data transfer agreements
Data Act	0	3	1	7	2
DGA	2	1	0	6	3
Freedom of Access	0	1	0	1	2
Directive INSPIRE	0	1	0	1	2

Source: own elaboration.

TABLE 4.

Comparison of disincentives based on frequency.

Legislation	Legal uncertainty	Manufacturer technical control	Transaction and compliance cost
Data Act	5	3	4
DGA	6	0	10
Freedom of Access	0	0	0
Directive INSPIRE	0	0	0

Source: own elaboration.

2.4.2 Regulatory incentives for accessing and sharing environmentally relevant data in the GDDS

2.4.2.1 Neutrality duty and structural separation

The DGA introduces two key entities: Data Intermediation Service Providers (DISPs) and Recognised Data Altruism Organisations (RDAOs), outlined in Arts. 10-15 and Arts. 16-25, which offer indirect incentives for stakeholders to engage in data sharing. For example, the DGA, Art. 12(a) states that they *shall not use the data for which they provide data intermediation services* ***for purposes other than to put them at the disposal of data users*** *and shall provide data intermediation services* ***through a separate legal person***.

DISPs act as neutral third parties, tasked with building trust and fostering collaboration among data holders, data subjects, and data users. Article 2(11)(a) DGA excludes certain services from becoming DISPs, such as data consultancies or providers of data products.

The neutrality duty for DISPs is expected to serve as direct incentive encouraging stakeholders to participate in data exchanges, reassured by DISPs' commitment to facilitating seamless data sharing while upholding rigorous data protection standards. Additionally, DISPs are poised to address longstanding barriers to data exchange, such as information asymmetries and regulatory burdens associated with data processing under data protection law, thus promoting more efficient data transactions.

DISPs operate as entities tasked with enhancing trust in data sharing and addressing information asymmetries by acting as impartial intermediaries between data holders and users, they are anticipated to offer stakeholders – e.g. environmental agencies, researchers, and private sector entities – the assurance of lawful and fair, reasonable, and non-discriminatory (FRAND) data use. Furthermore, **DISPs incentivise stakeholders through various economic and technological benefits**. For example, they can facilitate suitable matches between data holders, data subjects, and data users, ensuring that data are exchanged efficiently, effectively and in an interoperable manner.

RDAOs, conversely, are required to function as non-profit entities, legally distinct from any for-profit organisations, and conduct data altruism endeavours through a framework that is operationally separated from their other functions. By virtue of their legal independence, RDAOs are expected to **incentivise a culture of data sharing by encouraging voluntary contributions of environmentally relevant data without direct financial compensation**. This approach should simplify access to and sharing of data for the public interest (e.g. data related to air quality monitoring, biodiversity mapping, climate change mapping, and natural disaster preparedness), further facilitating collaboration, innovation, and public engagement within the GDDS.

The DGA's regimes for DISPs and RDAOs may help reduce various types of transaction costs, including ex-ante (pre-contractual agreement) and ex-post (post-contractual agreement). Ex-ante transactions relate to establishing suitable data sharing arrangements, and include, e.g. investments or expenditures made in preparation for a specific transaction, such as technology upgrades or specialised training, and costs incurred in making decisions regarding the terms and conditions of a transaction, such as the time and resources spent on evaluating options. Ex-post transaction costs can include monitoring, enforcement, and adjustment costs associated with ensuring that a contract's terms are met.

DISPs can reduce transaction costs for data holders and users in B2B and C2B data sharing, e.g. through standardisation and technical and contractual management of data transfers and enforcement of the agreed conditions (Richter, 2023). Additionally, RDAOs are expected to reduce transaction costs by fostering a culture of cooperation among stakeholders who can contribute data without direct financial compensation, thereby making it easier to access and share data for common purposes.

The services offered by DISPs and RDAOs not only have the potential to lower transaction costs and administrative burdens but also to enhance transparency and interoperability.

This fosters an environment conducive to data sharing and collaboration. In addition, by addressing concerns surrounding privacy, security, and legal compliance, DISPs are expected to unlock the potential for data sharing and reuse within the GDDS and across various common European data spaces.

2.4.2.2 Financial compensation

The legislation addresses this incentive in a direct way. To incentivise the continued investment in generating and making available valuable data, including investments in relevant technical tools, the Data Act contains the principle that data holders have the right to request reasonable compensation, which may include a margin, when legally obliged to share data with recipients in B2B relations. For example, Art. 9(2) and (3) states that *when agreeing on any compensation, the data holder and the data recipient shall take into account in particular: (a)* ***costs incurred in making the data available****, including, in particular, the costs necessary for the formatting of data, dissemination via electronic means and storage; (b)* ***investments in the collection and production of data****, where applicable, taking into account whether other parties contributed to obtaining, generating or collecting the data in question.* Later, Art. 20 stipulates compensation in cases of an exceptional need in case of B2G data sharing, *the data holder shall be entitled to* ***fair compensation for making data available*** *in compliance with a request made pursuant to Article 15(1), point (b). Such compensation shall cover the technical and organisational costs incurred to comply with the request including, where applicable, the costs of anonymisation, pseudonymisation, aggregation and of technical adaptation, and a reasonable margin.*

It is important to note that these provisions are not intended to purchase the data itself; rather, they enable data holders to receive fair compensation for the act of sharing data or, in the case of micro, small, or medium-sized enterprises and research organisations using the data on a not-for-profit basis, for covering the direct costs and investments associated with data sharing. The Commission should develop guidance detailing what qualifies as a reasonable compensation in the data economy (Data Act, Art. 9(5)). Such reasonable compensation may include firstly the costs incurred and, except for micro and small enterprises, investment required for making the data available.

Fees can also be charged by public authorities for, respectively, sharing spatial datasets and services with other public authorities, and for supplying environmental information, in accordance with INSPIRE and the Freedom of Access Directive. INSPIRE, Art. 17, states that *Member States may allow public authorities that supply spatial data sets and services to license them to,* ***and/or require payment from****, the public authorities or institutions and bodies of the Community that use these spatial data sets and services.* The Freedom of Access Directive, Art. 5, stipulates that *public authorities* ***may make a charge for supplying any environmental information*** *but such charge shall not exceed a reasonable amount.*

As set down in the DGA, public bodies can also charge fees but *shall take measures to provide incentives for the re-use of the categories of data referred to in Article 3(1) for non-commercial purposes, such as scientific research purposes, and by SMEs and startups in accordance with State aid rules* (Art. 6.4).

2.4.2.3 Portability requirement

The Data Act, Art. 23, mandates that *providers of data processing services shall take the measures (...) to enable customers to switch to a data processing service, covering the same service type, which is provided by a different provider of data processing services, or to on-premises ICT infrastructure, or, where relevant, to use several providers of data processing services at the same time*. Data holders, i.e. typically manufacturers of connected products and providers of related services, must offer customers the option to switch to a comparable service provided by a different data processing service provider. Within the GDDS, examples of connected products can include IoT smart energy meters, IoT smart agricultural devices, and environmental sensors, among others.

The obligation provided by the Data Act supplements the right to data portability outlined in Art. 20 of the General Data Protection Regulation (GDPR). Providers of data processing services *shall not impose and shall remove pre-commercial, commercial, technical, contractual and organisational obstacles* (Art. 23), which hinder customers from terminating contracts, entering new agreements, transferring exportable data, and achieving functional equivalence when transitioning to a new service provider. For example, concerned entities under the Data Act are required to facilitate customer data transfers within a 30-day timeframe. This switching provision can be seen as a direct incentive to eliminate barriers to change, such as the absence of interoperability standards. Failure to comply will result in penalties imposed by Member States.

The ability to switch between providers of data processing services, as outlined in the Data Act, serves as a direct incentive. For example, it can discourage cloud vendor lock-in that hinders data sharing with external parties due to compatibility, interoperability, or security issues. Consequently, it provides data users with the flexibility to choose alternative service providers, helping to remove barriers to change and fostering competition among providers. This competition can lead to enhancements in products, services, pricing, and user experience, ultimately benefiting customers. Additionally, the threat of penalties for noncompliance with the Data Act serves as an indirect incentive for covered entities to facilitate data transfers and comply with the requirements outlined in the legislation.

2.4.2.4 Public security guarantees in data transfer agreements

Stakeholders have the right to restrict public access to environmentally relevant data to protect national defence and public security. For example, the Data Act, Art. 32(c) states that *the addressee of a decision or judgment may ask the opinion of the relevant national body or authority competent for international cooperation in legal matters, in order to determine whether (...)* ***the data requested concerns national security or defence interests of the Union or its Member States***.

While this measure can be regarded as a disincentive, we can argue that it may encourage having a robust data security agreement in place, which incentivises organisations to share data by providing them with confidence that environmental information deemed as sensitive – for example, certain geolocation data – will be adequately protected from unauthorised disclosure. Accordingly, this right can be deemed as an indirect incentive, providing

flexibility in deciding whether to share data and what data may pose significant risks. Stakeholders include users and data holders (Data Act, Art. 4(2)), public sector bodies (DGA, Art. 3), and Member States (INSPIRE, Art. 13; Freedom of Access, Art. 4).

However, it is important to emphasise that additional measures – such as legal agreements or technological solutions – are required for this aspect to function as an incentive. As mentioned above, data security agreements outlining specific measures and protocols for protecting data can help mitigate the risks associated with sensitive, environmentally relevant data sharing. When organisations know that there are clear guidelines and procedures in place to safeguard sensitive information, they may feel more comfortable sharing data with other parties. A well-crafted data security agreement can serve as a foundation for fostering a culture of responsible, environmentally relevant data sharing while providing organisations with the assurance they need to engage in collaborative efforts without compromising sensitive information. In addition, some technological approaches can help in overcoming privacy and security concerns. For example, data visiting (e.g. for AI) can be used as an alternative to data sharing, where instead of transferring data across organisations or borders, algorithms or models are brought to the data source location. This practice allows for data analysis and, e.g. machine learning model training, without the data leaving their environment.

2.4.2.5 Commercial guarantees in data transfer agreements

All the four data laws give stakeholders the right to limit public access to competitively sensitive environmentally relevant data, particularly within the private sector. For example: the Data Act, Art. 19, states that B2G data requests shall: *(b) have implemented* ***technical and organisational measures that preserve the confidentiality and integrity of the requested data and the security of the data transfers****, in particular personal data, and safeguard the rights and freedoms of data subjects*. Another example from Art. 19 states, data requests ***shall not: (a) use the data or insights about the economic situation, assets and production or operation methods of the data holder*** *to develop or enhance a connected product or related service that competes with the connected product or related service of the data holder.*

Analogous to public security, this is a dual-edged factor, which can act as a disincentive in some situations, but also incentivise data sharing by providing flexibility in determining which data to share and underscoring the importance of protecting legitimate economic interests. In our analysis we considered this right as an indirect incentive when stakeholders – especially the private sector – perceive restricting public access to environmentally relevant data as beneficial for achieving certain goals, such as maintaining a competitive advantage or protecting proprietary know-how and *any other information the undue disclosure of which would have an impact on the market position or financial health of the undertaking* (DGA, Recital 10).

Public security and commercial confidentiality both serve as indirect incentives, as the inability to protect these interests may lead data holders to hesitate or refuse sharing data altogether.

2.4.3 Regulatory disincentives for accessing and sharing environmentally relevant data in the GDDS

2.4.3.1 Transaction and compliance costs

Transaction and compliance costs can apply to most of the actors, from data intermediaries (who are now mandated to register, which in many cases results in administrative and legal expenses to existing companies) to commercial data holders of connected devices, and public bodies mandated to allow the reuse of their data. This disincentive was observed in several passages in the DGA and likewise in the Data Act (Table 4).

Consequently, actors may hesitate to invest in or expand operations to enforce reuse conditions to 'ensure that the protected nature of data is preserved', as Art. 5(3) of the DGA mandates, necessitating legal advice and resource allocation for compliance measures to mitigate potential risks. The associated compliance costs can be seen as a direct disincentive, especially for entities with limited resources. However, this can be viewed as an overall trade-off when introducing new legislation, emphasising the importance of finding a balanced approach.

As previously mentioned, DISPs can streamline data sharing and reuse processes through their platforms, protocols, and frameworks. If data holders or subjects decide to use DISP services, data transactions are routed through these entities, thereby introducing a layer of costs for clients. 'The commercial terms, including pricing, for the provision of data intermediation services' (Art. 12b, DGA) can thus introduce transaction costs for data holders and data subjects, possibly including fees charged by DISPs for their services, as well as administrative and operational expenses associated with managing data transactions. While it is true that this perspective may appear to focus primarily on the business and revenue models of DISPs, it should be considered whether **these costs may act as direct disincentives** for data holders and subjects. Although these costs are intended to sustain the operations and ensure the viability of DISPs, they may impose a financial burden that deters some clients from using DISPs, or prompts them to seek alternative solutions. Therefore, while business and revenue considerations are important, it is also crucial to evaluate how these transaction costs may affect participation and engagement with DISPs within the data ecosystem.

Regarding the Data Act and the portability requirement, Art. 25(1) mandates that the *rights of the customer and the obligations of the provider of data processing services in relation to switching between providers of such services or, where applicable, to an on-premises ICT infrastructure shall be clearly* **set out in a written contract**. The legal mandates come with obligations to provide pre-contractual information concerning the generated data. This includes specifying whether the data is continuous and real-time, the manufacturer's intentions regarding data usage – whether for themselves or third-party use – the intended purposes of data usage, identifying the data holder, and outlining the user's data access rights. The need for contracts between data holders and data users as required by the law, as well as with third parties, has faced criticism for its perceived lack of legal basis. Economists have highlighted that the resulting increase in transaction costs could render the Data Act 'ineffective' (Kerber, 2023). According to

Kerber (2022), it remains uncertain whether these transparency requirements restrict manufacturers and data holders in modifying generated data, sharing practices, and intended uses over time. In a dynamic GDDS, there is a necessity for a certain level of adaptability concerning data generation and usage, especially for persistent products like IoT devices.

2.4.3.2 Legal uncertainty

We address **three aspects of legal uncertainty** which can serve as indirect disincentives. First, to comply with the DGA, RDAOs, which are based on principles of public interest as opposed to solely commercial motives, will encounter distinct legal inquiries concerning the application of data protection legislation within the realm of data altruism. Article 2(1)(a) mandates that RDAOs *shall inform data subjects or data holders prior to any processing of their data in a clear and easily comprehensible manner of: (a) the objectives of general interest and, if applicable, the specified, explicit and legitimate purpose for which personal data is to be processed, and for which it permits the processing of their data by a data user.* In particular, RDAOs must ensure that their activities align with the EU's Charter of Fundamental Rights and adhere to the GDPR's requirement of purpose limitation. This obligation entails clearly defining, specifying, and ensuring the legitimacy of the processing purpose. In this regard, Finck and Mueller (2023) see problems of compatibility with the GDPR. While the DGA assumes that data subjects' consent can legitimise the processing of personal data for altruistic purposes, the GDPR's requirements regarding valid consent – that it be freely given, specific, informed, and unambiguous, as well as withdrawable at any point – will be difficult to meet in practice.

Secondly, a contentious issue arises under Art. 15 of the Data Act regarding the requirement for private companies to share data upon request by public authorities for 'exceptional needs', which must be 'limited in time and scope' and may arise in response to, prevent, or aid in recovery from a public emergency (Recital 64). The term 'exceptional need' raises questions regarding its precise definition. According to Recital 64, 'exceptional needs' refer to unforeseeable circumstances that are limited in time, contrasting with planned, scheduled, periodic, or frequent events. Based on this definition, exceptional needs appear to pertain to discrete events such as natural disasters or accidents, which are typically sudden and localised. Unlike these events (i.e. natural disasters or accidents) environmental degradation and the effects of climate change unfold gradually over time and exert far-reaching and enduring impacts on ecosystems and human societies. Consequently, even though they may prompt emergencies or crises, they are recognised as ongoing, long-term processes within the broader realm of environmental concerns. Therefore, as observed by Finck and Mueller (2023), this lack of clear definition raises questions regarding whether public authorities can legitimately invoke the Data Act to access data aimed at addressing these prolonged phenomena.

The final point addresses the neutrality obligation upon DISPs under Art. 12 of the DGA. This Article stipulates that DISPs must operate in a strictly fiduciary manner regarding the data they handle. The processed data are tightly bound by specific purposes, with DISPs only permitted to use the data for the explicit purpose of making them available to data users (Art. 12 (a)). Any deviation from this purpose, such as using the data for alternative business endeavours, is expressly

prohibited. The rationale behind this stringent interpretation is believed to lie in the role of these provisions, which aim to generate trust in market activities and encourage data sharing among actors. However, according to Carovano and Finck (2023), economists caution that while the concept of neutrality is appealing in theory, there is a lack of empirical evidence to support whether its absence is a primary obstacle to data sharing through DISPs. In this respect, scholars such as Margoni et al. (2023) have proposed that limited data sharing could be linked to competition issues, such as the dominance of data holders and related information asymmetries, which the Data Act fails to tackle. Additionally, they highlight the necessity for coordination mechanisms to balance the desire on the part of data holders to safeguard trade secrets with the legitimate interests of users and selected third parties in accessing and utilising the data, an aspect not addressed by the DGA (Margoni, Ducuing, & Schirru, 2023). The implications for the GDDS in addressing competition concerns related to data sharing, information asymmetries, and control mechanisms can be significant.

2.4.3.3 Manufacturer technical control

Eckardt and Kerber (2024) argue that, historically, manufacturer technical control has been a barrier to data sharing, and that the Data Act aims at addressing this problem. One of the main objectives of the Data Act is to grant users of IoT devices and services greater access to the data generated by these devices, which is often controlled by manufacturers. The Data Act achieves this by introducing new rights for users to access and share IoT data with other entities. Article 3(1) of the Data Act mandates that *connected products shall be designed and manufactured, and related services shall be designed and provided, in such a manner that product data and related service data, ..., **are by default, easily, securely, free of charge, in a comprehensive, structured, commonly used and machine-readable format, and, where relevant and technically feasible, directly accessible** to the use.*

Granting users access to the data they produce (through connected devices and services) can act as a motivator for them to share it with third parties, be it for economic profit or societal and environmental advancement. However, the legal requirement presents significant implications and potential technical challenges for IoT device design. The effectiveness of the Data Act's attempt to incentivise data sharing by increasing data access rights beyond manufacturers seems unclear. Without complementary policy and legal actions, there is a risk that this intended incentive may not be realised and could inadvertently create barriers for data users, while turning into an indirect disincentive for IoT device manufacturers. For example, Eckardt and Kerber (2024) contend that manufacturers can opt for a design that allows them to retain exclusive control over the data generated by the device, which may be purchased, leased, or rented by firms or consumers. They argue that the Data Act does not challenge the manufacturers' strategy of exclusively capturing data via the technical design of IoT devices. Instead, it aims to mitigate the resulting adverse effects through two mechanisms: the establishment of non-negotiable user rights (Arts. 4 and 5) and the implementation of contractual agreements between data holders and users regarding the utilisation of non-personal IoT data (Art. 4(13) and (14)). To fulfil the goals of the Data Act, particularly in terms of making environmentally relevant data accessible for

innovation and public interest, Eckardt and Kerber propose that manufacturers **opt for 'bundles of rights' that reject exclusive ownership and facilitate widespread access to and sharing of IoT data**. They argue that the current Data Act, which remains overly committed to data exclusivity and imposes numerous obstacles to data sharing, is unlikely to significantly advance these aims.

2.4.4 Incentive trade-off – a scenario

The value of incentives for environmentally relevant data sharing can vary based on the goals, interests, and priorities of different stakeholders. For example, government agencies may focus on transparency and accountability, while private companies may prioritise innovation and product/service enhancement through data sharing. Interests also affect the willingness to share data, with researchers seeking diverse datasets for comprehensive analyses, and data subjects prioritising privacy and security. Stakeholders may also have different priorities when it comes to data sharing. For some, the primary concern may be ensuring fair and equitable access to data, while others may prioritise economic incentives or recognition for their contributions.

In Box 1, we present a hypothetical scenario describing a trade-off of incentives in B2G data sharing.

BOX 1.

Trade-off of incentives in B2G: The case of SolarCo.

SolarCo, a leading private company in the renewable energy sector, owns valuable data covering solar energy production efficiency, weather patterns, and consumer energy consumption. Therefore, SolarCo (the data holder) was contacted by an association of local authorities and regions (data users) interested in reusing the company's data to optimise energy distribution networks and offer tailored energy-saving solutions to consumers. Uncertain whether SolarCo would be willing to provide its data, the association hired a Data Intermediary Service Provider (DISP) to help negotiate while offering SolarCo an incentive to share (e.g. confidentiality guarantees to data transfer). In this respect, the neutrality duty of the DISP incentivised the municipalities, while the involvement of a neutral party in drafting the data sharing agreement bolstered SolarCo's trust. With commercial guarantees in place for data transfer agreements, SolarCo was encouraged to license its data to the association for a nominal fee, to cover administrative costs without overburdening the municipalities financially. Indeed, SolarCo recognised the critical role of sharing data with the public sector in advancing sustainability goals, such as encouraging renewable energy adoption and reducing carbon emissions.

However, SolarCo, as a leading company in the market of renewable energy, also holds an interest in commercially exploiting their data. This dual focus brings to light the role of the neutrality principle outlined in the Data Governance Act (DGA). In compliance with the DGA, Art. 12(a), the DISP ***cannot use the data for which they provide data intermediation services for purposes other than to put them at the disposal of the***

municipalities. Therefore, SolarCo cannot grant consent to the DISP for facilitating the commercial exploitation of their data, as this would conflict with the primary purpose of data sharing with the public sector. This obligation remains unalterable by mutual consent of the parties involved – SolarCo and the municipalities – unless specifically permitted by the DGA. Even if there were considerations to deviate from the data neutrality requirement for mutual benefit without compromising the common good (e.g. competition), any such agreement to deviate from this obligation would be legally null and void.

Although the neutrality duty provides a direct incentive for both SolarCo and the municipalities, it may result in additional costs for the DISP. The potential increase in costs for the DISP arises from the requirement to establish a legal separation of activities to comply with the neutrality duty. These additional costs could arise from various factors, including hiring legal experts to draft agreements and establish operational frameworks, investing in technology to separate and manage data appropriately, and dedicating personnel to oversee and manage the separated activities.

As a result, the DISP may need to allocate more resources and incur higher expenses to comply with the neutrality duty, potentially increasing the overall cost of providing intermediation services. These increased costs might render the DISP's services more costly for the municipalities, who are paying for the DISP services, potentially deterring their involvement in data sharing. Nonetheless, the collective nature of the request from an association representing multiple municipalities could yield economies of scale, aiding in offsetting the additional costs.

2.5 Conclusions and policy recommendations

Our examination has shed light on several regulatory incentives and disincentives potentially impacting data suppliers. From this analysis, we draw some reflections:

DEMAND-SIDE MEASURES

The analysis primarily examines the positive and negative incentives for data holders, though it may also be relevant to data users. Expanding the scope to encompass data users could provide a more comprehensive understanding of the dynamics at play within the GDDS ecosystem. Additionally, exploring how these incentives and disincentives manifest for both data holders and data users could offer valuable insights into the broader implications and potential solutions for a more efficient and equitable data space for all the stakeholders.

This study has not considered whether the legislation has enacted some demand-side measures. For example, incentives to match demand and supply, and/or to reduce transaction costs (search costs) to facilitate data exchange between supply and demand, or incentives to stimulate the demand for DISPs and RDAOs. In this respect, the examined legislation does not seem to explicitly provide incentives for the demand-

side of data sharing in the GDDS. As Ducuing (2024) also noted with reference to the Data Act and the DGA, there is a missing concern for the demand side. There is little discussion of how data sharing obligations can feed demand. Following Ducuid, the approach of the Data Act (and we add of the DGA) is based on the expectation that, provided they are well-functioning, markets will allocate resources (i.e. data) in an optimal manner (p.9).

Further measures, initiatives, or programmes might be necessary to encourage data sharing from the demand perspective, highlighting the associated advantages. These benefits include access to valuable insights, enhanced decision-making, innovation opportunities, and competitive advantages derived from using shared data.

APPROPRIATE CONDITIONS

Regulatory incentives for data sharing and reuse in the GDDS require appropriate conditions to be effective (Vitale, 2010). For example, appropriate conditions include trust, social capital, robust competition policies, and adequate technical infrastructure. Standardisation is also a crucial condition to enable data access and interoperability within the GDDS. Additionally, we can assume that incentives are contingent upon elements such as:

1. **Availability of internal capacity:** decisions by organisations regarding request of access, data sharing, and reuse may be influenced by the ability and efficiency of an organisation, often including the use of internal resources, such as technological infrastructure, expertise, and financial resources. Organisations with robust internal capacity are better positioned to participate in data sharing activities, while those with limited resources may face challenges that affect their ability to engage in data sharing effectively.
2. **Ease of appropriation:** this aspect relates to reinforcing mechanisms – included in the Data Act – which enable legitimate access for authorised uses, while preventing unauthorised access for potentially malicious or unethical purposes. The risk of not preventing unauthorised users can influence decisions by organisations regarding sharing of environmentally relevant data. As a result, individuals and organisations may choose to strategically share environmentally relevant data with trusted partners or stakeholders while safeguarding proprietary information through appropriate legal protections or trade secrets.
3. **Market size:** this variable can influence the willingness to share environmentally relevant data, particularly in industries where environmental concerns play a crucial role. The extent of market demand for eco-friendly products and services may impact decisions by organisations to share environmentally relevant data. For example, in larger markets with a substantial demand for environmentally sustainable solutions, organisations can be more incentivised to share environmentally relevant data. By doing so, they can demonstrate their commitment to sustainability, appeal to environmentally conscious consumers, and capitalise on commercial opportunities in the growing market for eco-friendly products and services.

In the light of the findings of this study, we recommend the following:

1. Recognise and address different stakeholder interests and priorities to develop incentive programmes that encourage a culture of data sharing while mitigating potential concerns.
2. Tailor incentive programmes to align with stakeholder priorities and needs. Consider types of incentives to promote standardisation for interoperability, stimulate data sharing, encourage innovation, and promote sustainability.
3. Consider new non-regulatory measures, such as voluntary participation programmes, industry self-regulation activities (e.g. data sharing guidelines and data sharing impact assessment), and data collaboration activities, to address disincentives. These initiatives can complement existing regulatory legislation and demonstrate industry commitment to upholding high standards for data sharing practices.
4. Consider pooling incentives. Pooling incentives involves aggregating resources or rewards and distributing them collectively among participants based on specific criteria or performance metrics. In the context of federated GDDS governance, pooling incentives can foster collaboration and collective action by offering shared benefits to individuals or organisations contributing to common environmental goals.
5. Consider regulatory incentives as a key attribute of a successful governance of a decentralised, interconnected, and transnational GDDS.
6. Consider experimenting with incentive structures in the GDDS using a sandbox for regulatory learning (example in Box 2). The lessons learnt from these experiments might inform new regulations and measures for data sharing; these measures might even be tested in a second phase of the sandbox.

BOX 2.

The GDDS Incentivisation Experiment: Exploring potential simulation scenarios using a sandbox for regulatory learning.

Objective

The GDDS Incentivisation Experiment aims to assess the effectiveness of various incentive structures in motivating data holders to share environmentally relevant data with the public sector. By simulating scenarios where data providers receive financial incentives for sharing data related to energy consumption, emissions, and resource usage, the experiment seeks to evaluate different incentive models and payout mechanisms within the GDDS.

EXPERIMENT DESIGN

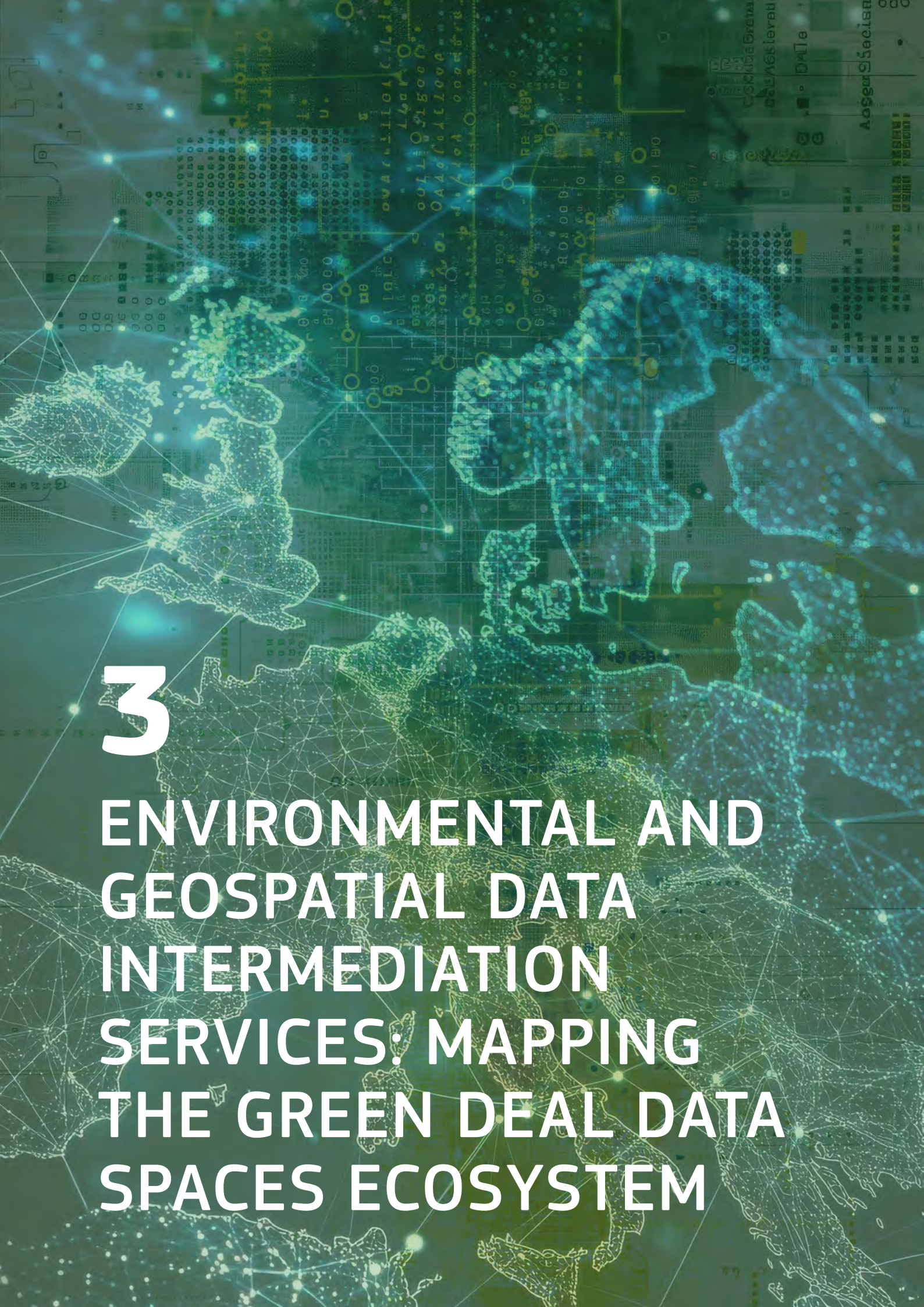
- **Scenario simulation:** Develop simulated scenarios that replicate real-world conditions for data sharing within the GDDS. These scenarios will vary incentive structures, payout mechanisms, and other relevant parameters to create a diverse testing environment.
- **Incentive structures:** Design and implement different incentive structures, such as subsidies, tax breaks, and other financial rewards, to incentivise data holders to participate

in the data space and share environmentally relevant data. Different data structures could be designed and implemented to test incentives on the demand side of data, encouraging individuals, businesses, and organisations to actively seek and use environmentally relevant data. For example, subsidies can be offered to incentivise data consumption and utilisation.

- **Payout mechanisms:** Experiment with various payout mechanisms to distribute incentives to data providers effectively. This may involve direct payments, tiered rewards based on data quality or quantity, or performance-based incentives.
- **Data collection and analysis:** Collect data on participation rates, data quality, and overall engagement of data providers throughout the experiment. Analyse the collected data to assess the impact and effectiveness of each incentive structure and payout mechanism.
- **Iterative refinement:** Based on the findings from the experiment, refine incentive models and payout mechanisms to optimise effectiveness in motivating data providers to share environmentally relevant data within the GDDS.

EXPECTED OUTCOMES

1. **Insights into incentive effectiveness:** Gain insights into which incentive structures and payout mechanisms are most effective in encouraging data holders to participate in the GDDS and share environmentally relevant data.
2. **Identification of best practices:** Identify best practices for incentivising data sharing in the context of environmental sustainability, which can inform future policy development and regulatory initiatives.
3. **Enhanced collaboration:** Foster collaboration between data holders, public sector entities, and regulatory bodies by creating a shared understanding of the benefits and challenges associated with incentivising data sharing.
4. **Contribution to GDDS objectives:** Contribute to the goals of the GDDS by facilitating access to comprehensive environmentally relevant data and supporting evidence-based decision-making and policy formulation.



3

ENVIRONMENTAL AND GEOSPATIAL DATA INTERMEDIATION SERVICES: MAPPING THE GREEN DEAL DATA SPACES ECOSYSTEM

3.1 Introduction

Using geospatial and environmental data to create value is a growing industry. In 2023, the European Association of Remote Sensing Companies identified around 770 companies in the sector. Since 2006, not only the number of companies grew, but the number of micro and small companies have particularly increased since the initial surveys. Data sharing in the context of the Green Deal is expected to increase in the following years along several sectors of the economy. The most prominent sectors are agriculture, environment, and forestry. With the new EU regulations that foster novel data governance schemes and the promotion of data exploitation, the INSPIRE goals should be updated to fit the needs of environmental data FAIRness[4] across stakeholders.

This research is focused on the implementation of the Data Governance Act (DGA) and its potential to foster an ecosystem of data intermediary services for Green Deal-relevant data (e.g. environmental, agricultural, energy and other sectors). Specifically, our research question is: **What is the role of data intermediaries in the upcoming Green Deal data space (GDDS)?**

The following is a non-exhaustive list of questions:

1. How can data intermediaries **instil trust and enable data sharing** in the context of the GDDS?
2. What are the current **roles and participation of data intermediaries** in the context of the GDDS?
3. What are the current **opportunities for finding value in data use** in the context of the GDDS?
4. What are the main **barriers for data intermediaries** to participate in data spaces?
5. What are the main **mechanisms for controlling and governing data** that are used in the context of GDDS?
6. What are the **data exploitation models** that would allow data intermediaries to interact in the context of GDDS?

To approach our analysis, first, we first define the context of the GDDS, and then present the categories of data intermediaries that can be found along this chapter.

3.1.1 The context of the GDDS

The Science for Policy Report 'Beyond INSPIRE. Perspectives on the legal foundation of the European Green Deal Data Space' (Kotsev, Escriu, and Minghini, 2023), published by the European Commission's Joint Research Centre (JRC), proposes different approaches to update the INSPIRE Directive. The report also highlights some of the current challenges of INSPIRE, which include:

- Outdated provider-centric legal framework with a strong focus on the public sector as the main user and provider of the data;
- Complex technical requirements that are enforced without an easily and objectively quantifiable benefit;
- Different trends and infrastructures being used on the national level in parallel with those put in place for complying with the requirements on the EU level;
- Novel technological developments and inclusion of new actors in the data economy (such as data intermediaries) that are not yet fully exploited.

On the other hand, the JRC Policy Report 'A Public Sector Contribution to the European

4. Findability, accessibility, interoperability, and reusability.

Green Deal Data Space' (Cetl et al., 2021) concludes that there is a need for evolving from complex and highly specialised geospatial data frameworks, where legal obligations are enforced by strict technical specifications, to flexible, open, agile, and self-sustainable data ecosystems. This is meant to be done through the GDDS, where different societal actors and communities can adopt INSPIRE tools, standards, and technologies to participate in the data market according to their own needs and requirements. This approach sums-up with the conclusions from the technical report on 'Emerging approaches to data-driven innovation in Europe' (Granell et al., 2022) that consider different lessons learnt from existent technologies in the context of the data ecosystem of smart cities.

The novel European regulations, the Data Act and the Data Governance Act (DGA), define the new rules to foster new ways of data sharing. This regulatory framework offers new opportunities to increase participation in the data economy, where communities, private companies and other organisations can benefit from the value created by new data sharing practices.

3.1.2 Discussing data intermediaries and intermediation relevant entities

This study looks at existing data intermediaries and intermediation-relevant entities in the environmental and geospatial domains. While our analysis conceptually falls under the logic of the DGA, it does not aim to determine legal requirements, definitions, or classifications of data intermediaries under such regulation. Rather, our intention is to study relevant practices around data intermediation to inform lessons and recommendations for policymakers and practitioners.

With this chapter we found that, to support the ecosystem within the context of GDDS, many actors should be taken into account. For this reason, we focus on the following data intermediary and intermediation-relevant entities:

- **Data intermediaries:** those that act as providing mechanisms for exchanging data between data holders/ subjects and data users;
- **Service providers:** those that provide additional services relevant to accessing, sharing, and using data, including data processing, and;
- **Technology enablers:** those providers of technology that facilitate the data sharing and exchange to third parties.

We consider **data intermediaries (DIs) as entities that enable and/or facilitate data sharing between data holders and data users, and are central players to leverage fair and inclusive data sharing practices**. In accordance, data intermediation activities can encompass a broad range of concepts. The JRC report 'Mapping the landscape of data intermediaries' provides further details on the roles and functions that DIs may perform in the context of inclusive data governance:

> *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.* (Micheli et al., 2023, p.11).

Some of the key recommendations provided in the JRC mapping exercise are the development of sustainable business models, definition of data privacy principles, and the governance models that allow these data intermediaries to work on behalf of their beneficiaries. The report defines 'control and agency', as well as 'value and benefit sharing' as some of the main challenges for data intermediaries.

The report identifies six types of data intermediaries: personal information management systems (PIMS), data sharing pools, data marketplaces, data cooperatives, data trusts, and data unions. Taking into account the entities found in the JRC report and those found in this chapter, we can see several coincidences and similarities. We view this comparison as a complementary work to support the findings in the report. Since we focused mainly on the context of the GDDS, we observed more service-oriented entities than data intermediation-focused ones, while the JRC report excluded service providers beyond data intermediation strictly. In particular, some other prominent sectors present more data intermediaries, in particular those focused on private and personal data (e.g. healthcare, finance). Not all the entities were examined under the same criteria. JoinData, for example, was considered a PIMS with the exception that no personal data was considered.

As a result, our research embraces different types of entities that are relevant to data intermediation in a broad sense, both within and beyond the definitions encapsulated in the DGA. In particular, the introduction of the **intermediation-relevant entities** completes the ecosystem of necessary actors in the context of the GDDS. We seek to contribute to clarifying some of the needs and main questions concerning the establishment of the GDDS by exploring the ecosystem of Green Deal-related data intermediaries and comparing them with the definition of data intermediation service providers (DISP) defined under the Arts. 10 and 12 of the DGA (European Union, 2022).

3.1.3 Data intermediaries in the context of the Data Governance Act

While data intermediaries can take multiple forms and functions, the DGA differentiates two main entities within this context: **Data intermediation services providers recognised in the Union (DISPs)** and **Data altruism organisations recognised in the Union (RDAOs)**. In a simplified way, DISPs are defined as entities that establish commercial relationships for the purposes of data sharing between data subjects/holders and data users. In the EU, those entities whose definition aligns with the DGA, have the obligation to notify national competent authorities about their activities and register as DISPs. They can be platforms or databases enabling the exchange or joint use of data, personal data management systems, and data cooperatives. In all cases, these entities should be neutral, meaning that there must be a structural separation between the intermediation services and other services provided to ensure they do not extract any direct profit from the information it is shared. RDAOs, are instead, non-profit organisations that share data for general interest objectives. Differently from DISPs, their registration is voluntary. Article 2 of the DGA defines data altruism 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.*

While RDAOs could play an important role in the context of data intermediation,

the present chapter limits its focus to the commercial approach to data intermediation (rather than an altruistic one) including DISPs and other intermediation-relevant entities. Correspondingly, Chapter 4 of this report provides an analysis of RDAOs in the context of the GDDS.

Based on the definitions provided by the DGA and additional existing studies on the field of data intermediation, we highlight four key concepts that are relevant for our research:

TABLE 5.

Summary of relevant concepts from the DGA that apply to data intermediaries.

Commercial relationship	A DI aims to establish commercial relationships for the purposes of data sharing.
Purpose and limitations of data use	Limits on the use of accessed or shared data for different purposes rather than for the data exchange.
Additional services	Limits on the possibility to aggregate, enrich or transform the data for the purpose of adding substantial value.
Structural separation	Additional services should be provided through a separate legal person.

Source: own elaboration.

Commercial relationship. A key criteria that should be taken into account in differentiating other data intermediaries is found in the Art. 2 of the DGA. This passage 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))

Purpose and limitations of data use. We point out the requirement established in the Art. 12, 'conditions for providing data intermediation services', which limits, among other things, the use of accessed or shared data for different purposes rather than for the data exchange. In particular, for the GDDS, with the goal of providing a social benefit or efforts towards a public interest, much of the data are openly accessible, and the economic return for data sharing might not be sufficient for these entities to operate. Thus, we propose to adopt a broad definition for this analysis. The limitations on the purposes for which a data intermediary may use data mentioned above were discussed by Richter (2023) who points out ambiguities in the DGA, such as the fact that entities might be considered data intermediation services 'only if they aim to establish commercial activities with regard to data sharing' or the fact that 'the data intermediation services may not use data for which it provides its intermediation services 'for other purposes than to put them at the disposal of data users'. (Richter, 2023, p.463). The interpretation of this rule has been controversial and for the purpose of this research, we take into account all evidence related to bringing value to data sharing and to establish relationships based

on the exchange of data, not necessarily as a financial transaction.

Additional services. In addition, some services are explicitly excluded in the **DGA**, such as Art. 2(11)(a) *services that obtain data from data holders and 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, without establishing a commercial relationship between data holders and data users*. Other exclusions pertain to copyright data, public services, closed groups or single holders of data. Moreover, in Recital 28, the provision of cloud storage, analytics, data sharing software, web browsers, browser plug-ins or email services should not be considered to be data intermediation services with the justification that *such services only provide technical tools for data subjects or data holders to share data with others, but the provision of such tools neither aims to establish a commercial relationship between data holders and data users nor allows the data intermediation services provider to acquire information on the establishment of commercial relationships for the purposes of data sharing*. Is not our intention to discuss the DGA or explore its limits. **However, even when we agree that those technical tools alone are not enough to consider them as data intermediation services, they are instrumental in making data available in a usable form and could expand the universe of entities informally considered as intermediation-relevant entities**.

Structural separation. In order to ensure **neutrality** in the services provided, the DGA mandates data intermediaries to establish a legal separation when these entities aim to provide additional data uses (and services related to the data) (Art. 12(a)). As pointed out by Ritcher, 'This separation principle aims to prevent conflicts of interest with and limit the risk of cross-data usage' (2023, p.463). However, this separation does not apply when the additional services provided are oriented to the facilitation of data sharing among parties, for example, 'temporary storage, curation, conversion, anonymisation, pseudonymisation', for the purposes of data intermediation services (Art. 12(e)).

3.1.4 Thematic data in the context of the GDDS

The GDDS is considered as an intersecting environment where different types of data could be shared. The nature of this approach is because the GDDS is a thematic and goal-oriented data space (seeking to solve global challenges, such as the Biodiversity Strategy, Zero Pollution Action Plan, Climate Change Adaptation Strategy), rather than as with sectorial intermediaries (e.g. those focusing on vertical markets, such as energy or mobility sectors).

Following the previous analysis of the European Commission on the INSPIRE data themes and high-value datasets, there are specific types of data considered critical for the GDDS:

- Geospatial
- Earth observation and environment
- Meteorological
- Statistics
- Companies' environmental sustainability performance data
- Mobility

Some examples identified by GREAT (Green Deal Data Space Foundation and its Community of Practice) project (2024) on how these data can contribute to use cases include:

- Local data about water reservoirs (gauge data, temperature, streamflow).
- Near real time in situ meteorological observations in a unified cross-border specification.
- Near real-time electric energy consumption of regions and the live CO2 footprint.
- High-resolution gridded information on biodiversity.
- Unified species occurrence data, including citizen science data.
- Indication of data set quality level and quality control.
- Socio-economic statistics that are not reported standardised[5].

Therefore, it is important to acknowledge that data should not be limited to those datasets but oriented to contribute to the achievements of such goals (e.g. zero pollution, a circular economy, biodiversity, or climate adaptation[6]). For example, data in the public sector can be used to study the state of the environment (*environmental awareness*), to achieve environmental goals or to evaluate and monitor implemented policies (data related to the environment); but, in the private sector the uses of data are much broader (Finck and Mueller, 2023). In addition, it should be taken into account that data intermediaries add new value to datasets, and this can be done by processing data from different sources that are not listed as strictly environmental or geospatial. For example, the following topics can be grouped within the industrial and commercial-related themes:

- Industry
- Energy
- Waste management
- Logistics
- Resilience

3.2 Methodology

3.2.1 Objectives

The main objective of this study is to understand the role of data intermediaries and other intermediation-relevant entities in the context of data spaces that can contribute to the exchange of geospatial and environmental data. This aim is carried out as a system thinking approach and is twofold:

Objective 1: The research consists of **mapping the actors in the ecosystem related to environmental and geospatial data** and understanding their activities in a future GDDS in Europe. With a focus on value creation of data, the objective is to map the ecology data intermediaries (of entities and actors, together with associated opportunities) for environmental data sharing and the GDDS. The goal is to understand how different actors and their motivations drive the negotiation on how the structures of data spaces are formed.

Objective 2: The current development of data spaces around different sectors (e.g. mobility, energy, smart and sustainable communities and cities) reveals several **interconnections (differences and similarities) between these data spaces in the context of the Green Deal**. Therefore, the present study conducts a deeper analysis of the interconnections between various data space

5. Green Deal data can help to identify inequalities and to describe socio-economic statistics as proxies by providing additional methods. For example, energy consumption or water quality analysis could provide insights to detect sectors with energy poverty or water deprivation. These can work as income inequality and migration and integration statistics, however, these should be proved and analysed in a case by case basis.

6. The European Green Deal https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2019%3A640%3AFIN

ecosystems and geospatial and environmental data intermediation services.

3.2.2 Data collection and validation

The research consisted of desk research through document analysis from grey literature collected using the snowball sampling method exploring multiple websites, documentation, and public material from relevant actors. The literature gathered was validated through informal communication exchanges with different actors in the ecosystem. Documents that have been collected are different in nature and can be classified in three different groups, serving our objectives:

- **Group A (for objective 1):** Communication material of actors in different ecosystems that could act as data intermediaries:
  - Actors involved in preparatory actions that contribute with different know-how and expertise on data space development.
  - Actors that actively participate in the environmental data ecosystem.
- **Group B (for objective 2):** Different EU funded preparatory actions for data spaces (e.g. smart cities, mobility, energy, etc.) have been funded during the last three years. Data spaces preparatory actions published specific documentation and blueprints for data space development in the future. This documentation includes standards and governance schemes that will overlap on the goals and specifications for the GDDS.
- **Group C (for objective 2):** Other organisations related to data standardisation (e.g. FIWARE, W3C, OGC) and data spaces (e.g. IDSA, BDVA) have released several documents on how to standardise spatial and environmental data exchange.

3.2.3 Data analysis

From the collected data and in concordance with the two objectives (see section 3.2.1), we performed a qualitative thematic analysis setting up two different dimensions of analysis.

For objective 1, we analysed the differential aspects of data intermediaries and other entities based on a set of inquiries related to their data sharing activities and business model. We defined three dimensions of analysis: technical (data and interoperability mechanisms), organisational (typologies of actors based on their business case and governance model), and value (context and added value to be exploited)[7]. These dimensions allowed us to define a set of questions for inquiries. In addition, we analyse the data theme or sector that each data intermediary approaches.

We defined the following specific questions as it follows:

- Data: What type of data are used?
- Interoperability: What data are accessed or shared and under which standards?
- Business case: How is value created and distributed among actors (holders/users/intermediaries)?
- Governance: Which type of participation arrangements are in place?
- Context: What is the business context?
- Added value: What is the key service that generate values?

7. We are mainly inspired by the Data Cooperation Canvas. The DCC was developed as in part of the preparatory actions for the Data Space for Smart and Sustainable Cities and Communities of the European Commission, to describe an existing data cooperation or to explore potential new cooperations within data spaces. This was a business tool to explore potential value models and not to develop an exhaustive research.

As a result, we obtained information from several entities that were classified in different aspects in our analysis.

For objective 2, we analysed the ecosystem of interrelated data spaces according to different dimensions: architecture, governance, interoperability, and data standards (including spatial).

Our definition of DIs is intentionally broad and includes all entities (private or public) that offer any data intermediation service for commercial purposes, including data analysis, visualisation or any interface meant to access the data. After our analysis, we compared our main findings to the DGA regulation, and annotated recommendations on how to apply the definition of DIs within the GDDS.

3.3 Mapping data intermediaries

When looking at the characteristics of the data ecosystem, we found that some of the requirements of the DGA – i.e. services based on a commercial relationship, the nature of the services provided, and the neutrality requirement – limit our ability to identify the broad group of entities playing a relevant role to facilitate data sharing within environmental and geospatial data ecosystems. Therefore, we have developed a categorisation that includes both the definitions of data intermediary service providers within the DGA, and other intermediation-relevant entities. It aims to structure an overview of the intermediation-relevant ecosystem in the GDDS by providing guidance on the existing governance structures and common business models and value creation strategies.

TABLE 6.

Categories of data intermediaries and intermediation relevant entities for categories falling outside the DGA.

Categories	Governance			Services provided				
	Single entity	Multiple entities	Federated	Data intermediation	Data collection mechanisms	Data visualisation	Aggregated and enriched data	Cleaning and anonymisation
Data collector	Yes	Yes	No	No	Yes	Yes	Sometimes	Sometimes
Data service provider	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Data provider (e.g. data catalogue)	Yes	Yes	No	No	No	Sometimes	No	No
Data intermediate through processing	Yes	Yes	No	Yes	Sometimes	Yes	No	Yes

Categories	Governance			Services provided				
	Single entity	Multiple entities	Federated	Data intermediation	Data collection mechanisms	Data visualisation	Aggregated and enriched data	Cleaning and anonymisation
Closed data intermediate	No	Yes	No	Yes	Sometimes	Sometimes	Sometimes	Yes
Data marketplace	Yes	Yes	Sometimes	Yes	No	Sometimes	No	No
Data space	No	Yes	Yes	Yes	Sometimes	No	No	Sometimes
Data cooperative	Sometimes	Yes	No	Yes	Sometimes	Sometimes	No	Sometimes

Source: own elaboration.

Data intermediation through processing. Some entities offer data processing in order to make data sharing possible. PylonData, North.io and DATACIE offer document scrapping and connectivity to data APIs to standardise and make this data available. They also offer insights and other data cleaning and quality evaluation processes.

Closed data intermediation. Some DIs might offer data sharing only to the members or stakeholders of their organisation, but not to external parties. Mercator Ocean could be a good example as they became an intergovernmental organisation where data are shared to their stakeholders who are, at the same time, actors that make data available by other means. Additional analysis would be needed to determine whether these entities fall within or beyond the scope of the DGA, in accordance with Art. 2(11)(c), which explains the exclusion of the following:

> *Services that are exclusively used by one data holder in order to enable the use of the data held by that data holder, or that are used by multiple legal persons in a closed group, including supplier or customer relationships or collaborations established by contract, in particular those that have as a main objective to ensure the functionalities of objects and devices connected to the Internet of Things.*

Data marketplaces. These are 'orchestrators of data sharing ecosystems that are open to all interested parties' (Recital 28 of the DGA). They expose and make it possible to access the datasets and usually provide mechanisms for downloading or accessing the data (offered by the same entity or another). An example of a marketplace could be SeaDataNet, which offers an advanced search engine for datasets at the same time it provides a visualisation preview of the data it offers.

Data spaces. Could be 'purpose- or sector-specific or cross-sectoral interoperable frameworks of common standards and practices to share or jointly process data' (Recital 27 of the DGA). These are not considered simple entities, but complex

governance structures that play a key role in the ecosystem for facilitating data exchange. Despite not being a DISP, we include them as part of our survey since they imply a more decentralised way of data sharing, differently from marketplaces. Even when data is listed, usually mechanisms for accessing data are made available through automated consent and more advanced standards for exchange. Typical data spaces within the GDDS context are AgriDataspace or the Copernicus data space.

Data cooperatives. Defined by the DGA as 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 …' (Art. 2(15)). Cooperatives have a clear objective referred to represent the members of the organisation towards data access and sharing. An example of data cooperatives in the context of agriculture is SAOS.

In the following subsections we present some exemptions and additional observations during our classification. In addition, **data collectors, data service providers, and data providers** (e.g. data catalogues) are further explained in Section 3.4 'Beyond data intermediation'. While they do not provide data intermediation services, they are considered as intermediation-relevant entities.

3.3.1.1 Data themes in the context of the GDDS

Following the data collection methodology, we have initially assessed and classified 51 entities. We found that the most frequent data themes are 'earth observation and environment' (19) and 'geospatial' (14), followed by 'agriculture' (7), 'industry' (6) and 'resilience' (5). While 'circular economy', 'mobility' and 'biodiversity' are the less present themes. Some entities work with more than one data theme (e.g. the Urban Data Big Data Centre works on mobility and geospatial themes). The most extended use cases are related to geographical information systems and geospatial analysis in general, although there were also more specific ones that focus only on the analysis of oceanic or hydrological data. Making satellite data available is the most common case of data sharing in those categories.

In continuation of our analysis, and in concordance with the analysis on data intermediaries, the most prominent themes in the data spaces analysed are the 'earth observation and environment' as well as 'geospatial'. 'Green Deal' appears as a theme on its own, but sometimes approaching only a subset of themes (see the GREAT project).

Bearing in mind that the Green Deal considers the impact of climate on biodiversity, social relations, job markets, and the global economy, if we look at data providers, new themes, such as 'biodiversity' or 'sociology and humanities' come to the fore. This puts into question the boundaries of the scope of data that the GDDS should consider. The flexibility in the criteria of categories should correspond to the goal-oriented logic of the green deal. Thus, it is worth asking, whether this data could be used for measuring and creating value towards a better planetary wellbeing?

Spatial and environmental data have their particularities. For example, emission data used to be taken from satellite imagery or satellite sensors, which are managed by public and/or private entities and require a special type of infrastructure to hold and process this data. Cetl et al. (2021) argue that spatial data infrastructures (SDIs) are

particularly challenging to maintain while at the same time ensuring privacy and security. Thus, these aspects may impose limitations in the business models of specialised data intermediaries in the context of the GDDS, who may be required to look for additional funding sources to manage the costs related to spatial data sharing. Since one of the DGA's goals is to provide neutrality and promote data accessibility, special attention would be needed to develop adequate business models for DIs of this type. It is important to have in mind that in Recital 27 of the DGA, DIs may include public sector bodies, which it makes it compatible to SDIs become a DISP.

3.3.1.2 Neutrality requirement and additional services

One of our main findings was that most of the analysed entities did not operate according to the conceptualisation of the DGA in a strict manner. This was particularly relevant when looking at the provision of data intermediation services together with additional data-driven services with a legal separation (i.e. neutrality requirement).

Some of them do offer particular services with the data obtained, such as emission reporting or weather forecasting. Agriculture-oriented services are intertwined with environmental regulations on the industry. Some other entities offer data-driven services related to logistics, production chains and the food supply, where most are focused on tracing products from their origin, offering waste management tools, or measuring greenhouse effects. A special category includes those that focus on risk caused by climate change, where data-driven analytics are offered to protect assets and prevent them from risks or provide data insights on the impacts of diverse production strategies.

When looking at those entities that solely provided data intermediation services, we also found that data intermediaries can form mutually beneficial relationships with other actors in the ecosystem, such as data visualisation entities. For example, JoinData provides data intermediation by offering farmers a data collection app to store and share their data. On the other hand, Farmdesk is a service provider entity that only offers data visualisation and reports to the same farmers. While JoinData charges an annual fee to farmers to use their services, Farmdesk can connect to JoinData and offer a differentiated product to make their data richer. These are different entities and are not associated at all, however, they complement each other in their offering.

Furthermore, we found that existing entities with a legal separation of their various services had complex structures formed by various associated entities. For example, PLACE is a transnational, multiple-entity data intermediator that makes data available for research through an NGO at the same time that it offers commercial licensing through a sister private entity. However, data are held by a data trust and licensed to their sister entities (e.g. Thisis PLACE Trading, ThisiPLACE Trust, ThisisPLACE UK, etc.). This complex structure allows them to make separate decisions about data holding, monetisation, and altruistic processes. Despite that PLACE can fit different descriptions depending on its unit, looking at their associated entities[8] we can consider the entity ThisisPLACE Trading as a data intermediary in the sense they commercialise the datasets, while ThisisPLACE Trust could be

8. PLACE associated entities, these include ThisisPLACE Foundation, ThisisPLACE UK, ThisisPLACE Trust and ThisisPLACE Trading. Accessed at: https://thisisplace.org/legal/

considered a RDAO and ThisisPlace UK offers additional data-driven services.

3.3.1.3 The ecosystem of data spaces and development stages

Data spaces could be defined as 'platform ecosystems built on federated infrastructures and participative governance structures for sovereign data exchange oriented toward stakeholders' interests' (Schurig, Kari, and Fürstenau, 2024, p.7). This novel perspective fosters decentralisation and focuses on neutrality of the network, but also on automated mechanisms for data access and sharing. However, this is not always the case in our research; many data spaces have less ambitions to develop a federated governance and seek for multilateral cooperation.

Therefore, our understanding is that data spaces can be considered relevant data ecosystems that facilitate data intermediation among various stakeholders. Differently from data intermediaries, we do not consider data spaces as entities in themselves (i.e. they are closer to the notion of a data ecosystem than to a data intermediation service provider). **In this subsection we reflect on the role of data spaces as data ecosystems that facilitate data exchanges across stakeholders, and provide an overview of their role for data intermediation.**

First, it is important to make clear that data intermediaries can also operate in the context of a data space (e.g. a data trust or a PIMS providing citizen data to a data space). This ambiguity is expected due to the complexity of the ecosystem. We found that some of the intermediaries identified in our analysis already participate in data space ecosystems (whether recognised data spaces[9] or not) and, from our observations, it seems a natural environment for intermediaries to appear. The question is, how we can help data intermediaries better serve data spaces by bringing trust and neutrality? To answer this, we analysed the ecosystems of data spaces around the themes identified in the context of the Green Deal.

Previous to data spaces, data marketplaces represented one of the main approaches to data intermediation, mostly organised as a centralised form of platform ecosystem (Van De Ven et al., 2021). Contrary to data spaces, we found that data marketplaces that have centralised management are more developed than decentralised and complex mechanisms, probably due to the novelty of the federated approaches.

Regarding their development stages, the International Data Space Association (IDSA) identifies in its Data Space Radar different maturity stages for data spaces: exploratory, preparatory, implementation, operational and scaling stages. We classified the data spaces under this criteria and we found that most of these are in the initial stages while only two out of 15 are in the operational or scaling stage.

Sometimes data spaces in any of the three initial stages might be indistinguishable from projects to enable data sharing due to the lack of documentation or evidence of their inner working (see Table 3 in Annex 1). In consequence, it might be that these projects' repositories are not maintained or updated. Therefore, this issue should be studied in depth. To date, it is yet unclear how many projects will evolve or change to advanced stages, becoming established data spaces with federated governance structures and up-to-date data.

9. Found in https://digital-strategy.ec.europa.eu/en/policies/data-spaces

3.3.1.4 Publicly funded data intermediaries

In the particular context of the GDDS, much of the data collection and data use has a public interest (e.g. from fighting climate change, taking care of biodiversity ecosystems, or to help the food chain). We found that these intermediates, even when they can be classified in the above categories, present some particularities in the context of the GDDS that are worth mentioning due to the type of business model based on public funding. We found three types of public influence:

1. Publicly owned data intermediaries that provide data intermediation services.
2. Many data sharing initiatives and intermediaries that have a business model that is mainly based on a public funding scheme with no clear business model that includes commercialisation of data. In these cases, we can consider those that might be beyond the scope of DISPs.
3. Data intermediaries that are publicly funded but in certain cases offer data use licences that should be taken into account as commercial relationships.

The implication of the presence of such type of intermediaries is that the GDDS presents a more complex scheme of entities. Some could be excluded because they might be considered as public services, whilst others might not be entirely DISPs because they do not aim to establish a commercial activity, rather, they could be considered *open data intermediaries* (Shaharudin, Van Loenen, and Janssen, 2023). In the case of research-oriented DIs, the DGA places an exception to be considered DISPs, but this does not prevent them from offering such services (European Commission, 2022c). We present examples of the three types identified above.

Type 1: A research institution acting as a data intermediary is the European Centre for Medium-Range Weather Forecasts (ECMWF). They provide meteorological data and services by using the Copernicus data. Their services can be used by different governmental offices, but they also provide services under commercial licences.

Type 2: The Global Biodiversity Information Facility (GBIF) is a publicly funded institution. It offers publicly available datasets under Creative Commons licences and excludes any licence that allows the exploitation of data for commercial use. Similarly, the European Planetary Observing System (EPOS) is a European Research Infrastructure Consortium (ERIC) in charge of providing data for research purposes. Although they accept private institutions as users of their services, it is unclear that results can be other than research.

Type 3: The Urban Big Data Centre (UBDC) is a publicly funded research institute in Scotland that offers free data for use, but some datasets have particular licences. In particular, specific uses of data should be requested on a case-by-case basis, also research purpose uses are allowed in data sharing free of charge. This is an ambiguous scenario where the institution might be considered a RDAO but also as an open data intermediary. The Biobanking and BioMolecular resources Research Infrastructure (BBMRI) is another ERIC, but it offers, on the one hand, a catalogue of data for which there are commercially available datasets, and on the other hand, it also offers a negotiator to connect and reach a consensus for data exchange. In this case, we consider that the entity is a kind of marketplace that could work both for non-profit and for-profit organisations.

3.4 Beyond data intermediation

As we have highlighted in this report, our findings point out that the ecosystem around the GDDS is much richer than expected. We argue that, to foster a data value economy, we need to look at the value generation processes and the actors involved in this process, which often include data processing and service provision. Following objective 2, we analyse how data spaces facilitate data access and sharing activities, through orchestrators, technology enablers, and other actors. Therefore, the importance of these entities is relevant to the creation of new synergies in the GDDS context. In this section we review some of the relevant entities and key processes that are part of the ecosystem and which generate the grounds of the appearance of data intermediaries.

3.4.1 Relevant intermediation entities

Data Collectors. Due to the limitations of developing profitable business models based solely on data intermediation, we found that several organisations in the Green Deal ecosystem focus rather on data collection activities, both quantitative and qualitative data. This is extremely important, because they can be considered data users with potential to become intermediaries. They offer different types of data visualisation and data access in non-interoperable ways. For example, SyncForce or Deltares offer access to the data they hold but only to specific clients. Deltares also provides access to visualise their data openly through its platform Blue Earth Data.

Data providers. These are entities that can offer solely a list of metadata related to specific datasets, without making these data accessible or shared through their catalogue. Sometimes, the data should be requested by a different entity such as the data owner or data holder. For example, the RUDI project from the city of Rennes (France) offers a catalogue of open data but also third party data that can be requested upon request.

(Data) service providers. Many entities offer data-driven services that are essential for other stakeholders in the ecosystem. In some cases, these entities also provide data intermediation services. For example, Creodias offers cloud computing as a service, while also facilitating access to satellite images and other environmental data. In another example, Circularise is a blockchain-based platform to track and trace products, allowing to measure the carbon footprint of industries. Such services include data sharing capabilities between data holders and data users in a granular manner.

3.4.2 Data standardisation and connectors

The main characteristics of data spaces reported in the first staff working document on data spaces (European Commission, 2022b) highlight the importance of securing the confidentiality and privacy to pool, access, share, process and use data; enable neutral and transparency forms of participation; and clear and trustworthy data governance mechanisms. To achieve this, preparatory actions for data spaces have been developing data standards, protocols and technologies that will be shared among the data space ecosystems.

Standardised data is needed for promoting the reuse of data. This is already well embraced in the Green Deal context for the majority of data. Most of the environmental and earth

observation data apply the Open Geospatial Consortium (OGC) standards and many others.

During 2022 and 2023 several initiatives were developing methods of standardisation and connection for data spaces. However, the current scenario is not free of challenges in the ecosystem. In the European Commission report second staff working document on data spaces (2024), a list of EU-financed data spaces was released, where only four Green Deal data spaces appear. The interoperability of geospatial and spatiotemporal data is more complex than other types of data.

3.4.3 Technology enablers

One of the characteristics of data sharing ecosystems is the need of infrastructure and technologies to enable interoperable capabilities. Some of these capabilities are developed by the same data intermediaries, but it is not always the case. We found that some providers offer technological support to create data intermediary services. This is of special importance to help the ecosystem to grow by promoting the inclusion of new entities into the data ecosystem. **Technology enablers are not data intermediaries per se, but offer the necessary tools and mechanisms to other entities to become interoperable**. This could be connectors, identity verification, standards, and other software. In Verstraete et al. (2023) a similar case is analysed, where for the 'organisation Z' authors consider that these intermediaries could be considered DIs but fall outside the scope of the DGA. The analysis of these entities falls outside of focus of this report, but readers can find the following examples: DesideDatum, Dawex, FIWARE, Nexyo or Tritom.

In the context of data spaces, these are called *data space intermediaries*[10]. They provide operational efficiencies and facilitate data transactions via logs or a clearing house. They can also help attract new parties to the data space. While a data space intermediary provides essential services to the data space, it is also a participant in the data space.

According to the definition from the Data Spaces Support Centre (DSSC), they can be distinguished by the type of technical service they offer and are categorised in three types:

- Service providers for enabling services and functions
- Connection-providing intermediaries
- Other data sharing mechanisms

For example, the Space Data Marketplace (which is a private data space), offers spatial data through Dawex data exchange technologies and connects a diversity of service providers and companies. Airbus uses the data for creating 3D digital twins, Thales Alenia makes a search engine, Geoflex offers last-mile logistics augmentation solutions, and Murmuration offers environmental monitoring services.

We have already highlighted the importance of the ERICs. These ERICs might have different governance structures and data policies, working as data spaces or as data intermediaries, but also as data space intermediaries. At the same time, each ERIC contributes to ENVRI-Hub, a joint data portal. Thus, the category in which an ERIC could fall depending on its governance structure.

10. Data Space intermediaries, available at: https://dssc.eu/space/BVE/367558657/Data+Space+Intermediary

3.5 Discussion

From our study we identified that most data intermediaries offer geospatial data and earth observation analysis. However, we accounted for a few intermediaries that already function as the DGA requires, some of them being public institutions or publicly funded initiatives (like ERICs). We can assume, as many authors have indicated (Micheli et al., 2023; Carovano and Finck, 2023; Richter, 2023), that finding a sustainable business model beyond public funding in compliance with the DGA is not easy. Nevertheless, it is yet early to prove this assumption.

Following our broad definition of data intermediation, the main topics of the data they work with relate to agriculture, water, and pollution. Some other intermediaries offer data relevant for logistics, mobility, or supply chain (e.g. enabling an industrial circular economy or ethical production) initiatives, while some offer weather forecast data. A special category for data intermediation services is asset management and risk management – these companies target large companies, financial, and insurance sectors.

Regarding data sharing, a few of them offer access to data through standards, but most offer specific services for exploiting data. Most of the earth observation data and geospatial analysis entities offer similar products. While some differentiate when focusing on particular industries or use-cases, others have broader targets. Most of the intermediaries analysed are SMEs (also certified B corporations); however, several non-profit are also considered part of the ecosystem since they offer commercial services by exploiting or sharing data. Some of them also participate in data science or environmental awareness campaigns, which is to be expected due to the importance of the topic for society.

3.5.1 Closing the gap between stakeholders by offering data value

As noted above, we should not limit our analysis to the type of data, but rather address the goals data intermediaries want to achieve, or at least help to achieve by instilling trust in the process of data access or sharing.

We acknowledge that data sharing as a business activity is closely related to other data-driven commercial activities and are very often provided in conjunction. As value is tightly linked among various activities in the data economy, the exploration of novel DGA-compliant sustainable business models would be beneficial in the GDDS. As data intermediation service providers, these entities draw from third-party data to offer particular services. In some cases, we found it difficult to distinguish whether the economic value was created in the intermediation of the initial data or in the service provided on top of it. In some cases, intermediaries could simply separate their business units and sign a framework agreement to continue providing services under a legal separation. However, it is important to note that a higher demand for data is expected in the upcoming years, where DISPs can play a key role helping to share data among more data users, fostering more value from the data they share in the ecosystem.

From the analysis, we learnt that data intermediation as a business model is more an exception than a rule. Since there are different providers competing in the same market they should differentiate their offers.

Within the spectrum of intermediation, it is sometimes difficult to classify these entities into specific categories. For example, some may facilitate data collection and data-driven services by extracting value from the collected

data, but not data sharing. Some others facilitate the tools for data collection only and make the data available without extracting value (becoming a data sharing platform). This might be the case of some actors that operate with citizen generated data, or grassroot data.

A particular type of intermediation-relevant case involves digital twins. Digital twins can be just projects or consortia created with different forms of organisations. In any case, these are complex projects that use pieces of software to convert a physical environment into a digital duplicate. Digital twins are of particular interest to the Green Deal since there are attempts to recreate rural and urban environments, for example, to analyse the impact of climate change and evaluate resilience strategies, or to analyse the effects of mobility. To create digital twins a vast amount of data is required, and should be exchanged and collected by different means (some may include data spaces). After training machine-learning algorithms, autonomous agents and other artificial intelligence mechanisms, the system offers visualisation and output data that can be used for decision-making. This is clearly a service that could be a use case or, considered a DI in itself since new data is created (often synthetic data). For example, Destination Earth is a digital twin project, funded by the European Union, that seeks to connect and share data, develop data analysis, and make new data available, becoming a notable use case. Digital twins can be seen outside the scope of the definition of DIs as they might be tools, infrastructure or mere data use-cases. But a different governance scheme could imply the creation of a data intermediary for this particular case.

Another complex scenario consists of the entities whose work involves cryptographic assets (Blockchain/web3) which act as technology enablers. In the decentralised nature of the blockchain, they do not hold data, nor do they govern the data. EnergyWeb, for example, is a non-profit organisation that facilitates energy data exchange, providing consulting services to implement their open-source technology. This could hardly be considered a DI, but also not a RDAO, a data space, or a marketplace in terms of data interoperability. Technology enablers are a kind of special category that also falls outside the scope of standardisation entities such as FIWARE or OGC. However, the application of blockchain can also be focused on offering a direct service, using the data to offer B2B or B2C products (e.g. OpenSC ensures fair and transparent food sustainability tracking. Even though this is a good example, it was not included in the entities analysed since it currently operates outside of Europe).

Data spaces, instead, are intended to offer opportunities to distribute and create value along their network. By using federated governance frameworks, they ensure trusted and recognised behaviour between stakeholders. We may look at the benefit these data spaces offer, not only in terms of their sectorial application but also cross-sectoral implementation. This means, the interoperability directed outside the data space could enhance the goals pursued by the Green Deal. For example, the energy sector data space and a mobility sector data space could contribute to the assessment of emissions in a food sector data space. Therefore, a GDDS could create value by accessing other data spaces but specifying the particular needs of data standards and procedures to accomplish Green Deal goals. An organic development could be expected if the availability of data and connectors are open to external actors, and a GDDS is orchestrated

to interact with other data spaces, while empowering communities and the users of data that might take advantage of this data.

3.6 Conclusions and policy recommendations

As we could draw from our findings, we consider a critical point to be the fact that actors in the data ecosystem can fall within many categories relevant to data intermediation. This is because definitions refer to different capacities and characteristics. Moreover, it is difficult to separate funding and goals of entities from their context (e.g. research institutions, public sector entities, public-private consortia, or citizen data initiatives).

From the perspective of the Green Deal, because of the public interest of the data and the goals pursued in this context, we explored many actors that are funded by public institutions and programmes. We found that many entities that provide data intermediation are closer to the RDAO definition or open data intermediaries to exchange data (most for research purposes).

1. Novel technological enablers should be considered as key actors in the GDDS ecosystem.

More interoperable features and an extended showcase of data would be expected of these entities to make a more attractive offering. We acknowledge the need for considering other types of intermediation-related entities (services, enablers, etc.) and not only data intermediaries in the regulation, because these are needed to generate more value in the ecosystem. These specialised entities could bring knowledge, technologies, and organisational tools that empower and facilitate the development of the data market, supporting the DISPs and RDAOs that provide data to sectoral or cross-sectoral use cases.

2. Data providers, service providers, and data intermediaries are often interwoven in a way that generate particular value for specific Green Deal goals.

Our findings show that many intermediaries are tightly interrelated in the context of the GDDS data themes. Those mainly intermediate industry-related data (e.g. supply chains, waste management, pollution, or risk-related products), use earth observation data and geospatial data, as well as mobility and energy data. In fact, the three use cases defined by the GREAT project (biodiversity, zero pollution, climate change) may not have specialised data intermediaries for each of them, but could be fed by many data sources. Thus, the complexity of actors relates to and is highly dependent on the use of specialised data that could benefit the role of the GDDS.

3. Additional incentives to leverage the opportunities created by the DGA should be explored in the near future.

The DGA was created to promote data sharing and to foster new types of collaborations through trust and neutrality. One of the mechanisms is the creation of a stamp of recognition (i.e. official logo), transparency reports and blueprints. However, it is yet unclear whether these mechanisms alone will generate the necessary demand for data sharing the GDDS. As such, additional incentives should be explored to complement the DGA and leverage long-term sustainable business models to support data sharing.

Furthermore, organic, flexible, and self-organised data spaces are essential to create new data sharing dynamics. Following the incentives, private data spaces might have

an interest in opening up to third parties, and interoperable mechanisms will be necessary. Deeper analysis should be done to perceive the value of data intermediation for each of the identified entities, and to analyse whether available resources and incentives (as seen in the previous chapter) are sufficient for them to adequately sustain their activities and respond to the needs of data holders and users.

4

DATA ALTRUISM AND THE GREEN DEAL DATA SPACE: MAPPING THE ECOSYSTEM

4.1 Introduction

The concept of 'data altruism' is first introduced by the European Commission in the Data Governance Act (European Union, 2022) (hereinafter DGA):

> *Data altruism' means the voluntary sharing of data on the basis of the consent of data subjects to process personal data pertaining to them, or permissions of data holders to allow the use of their nonpersonal data without seeking or receiving a reward that goes beyond compensation related to the costs that they incur where they make their data available for objectives of general interest (...).* Art. 16 of the DGA

Beyond public data, the DGA aims to increase the opportunities for other subjects (citizens, communities, private companies, and other organisations) to contribute to the data market, through new data sharing practices and for the benefit of advancing the general interest.

A key pillar of the European strategy for data (European Commission, 2020 and 2022a), the DGA has at its core the objective of increasing trust in data sharing through various means, including, for instance, the role of data intermediaries (cf. Chapter III). Whilst data intermediaries have so far received most of the attention from commentators, the DGA also introduces data altruism organisations, aiming to recognise and legislate the status of many organisations who work to enhance data sharing practices for non-profit and altruistic purposes.

Through the DGA, the Commission attempted to achieve two aims. On the one hand, to recognise the work of existing not-for-profit organisations that operate to enhance data sharing practices and that were not falling under the regime of data intermediation service providers. On the other hand, the Commission aimed to address the problem of trust on the part of private actors toward data sharing practices, by creating an EU Register for so-called Recognised Data Altruism Organisations (hereinafter RDAOs). Although registration is not compulsory for the purpose of data sharing, the assumption behind the establishment of these registers (both the EU and national[11]) is to provide criteria and standards to ensure the protection of the rights and interest of private actors. National registers, which are established and maintained by the national authorities of EU Member States, have the clear aim to increase trust amongst private actors in order to enhance data sharing practice with these recognised organisations[12]. The role and potential of RDAOs have so far remained mostly unexplored, especially with respect to environmental data debates, and this chapter aims to start filling this gap.

At the time of writing, the Rulebook, which should detail the technical requirements, the interoperability standards, and a communication roadmap for data altruism organisations, has not yet been published. In addition, national registers for RDAOs have also not yet been established in every Member State. Moreover, it should be noted that not all Member States have identified the national authorities who will be responsible for the creation and

11. The EU register of RDAOs has been established within the framework of the DGA. National registers are being established by the EU Member States. Once the national registers are in place, Member States will notify the Commission as and when data altruism organisations are registered. At the time of writing only one organisation is listed in the EU register: Datalog. More information on Datalog are available here: https://datalog.es/

12. The Commission has also introduced logos for RDAOs (as well as Data Intermediaries). This is in order to help actors easily identify data intermediation service providers and data altruism organisations that are recognised in the Union https://digital-strategy.ec.europa.eu/en/library/logos-data-intermediaries-and-data-altruism-organisations-recognised-union

maintenance of the registers (see Annex 1). Under these conditions, it is too early to provide a full analysis of the role and impact of RDAOs on EU data sharing practices, which would require further research in the future and once the registers for RDAOs would be in place.

The aim of this chapter is therefore to provide an initial and introductory examination of the potentials and challenges of data altruism models in general, and to sketch a future practical role and use cases for RDAOs.

4.1.1 The scope of the research: Key terms and debates

'Data altruism' debates and legal framing all depart from and carry with them a series of assumptions about key terms and their meanings. This section aims to provide a brief introduction to the concepts of 'altruism' and the concept of 'general interest', which are both central in the DGA and yet still contested.

'Altruism' is the word chosen by the legislators to describe this model of data sharing. The word altruism, it has been argued (Lalova-Spinks, Meszaros and Huys, 2023; Hansen et al, 2021), might have been used in preference to the term data donation. The latter in fact implies an ownership transfer that is not contemplated by the DGA and that might not be possible depending on the qualities and types of data possibly involved. In the case of environmental data, for instance, citizens might share data that they do not (and could not) own, including data that might involve other people (e.g. the data produced and collected within a household), or data that is generated by other actors and through not proprietary processes. Moreover, the notion of donation would imply that the party that donates the data no longer has access to it. When it comes to health data, this option would be unethical in most cases and illegal in some (Prainsack, 2019).

Whilst some scholars and commentators still wished for the EU to provide the legal ground for data *donation* as a possibility (Veil, 2021), most scholars have warned against the adoption of this framing. According to them, in the donation-based approach a proprietary-based framing of data is implied, that would not recognise, for instance, data as *co-constructed* (Ballantyne, 2020) and could be harmful to the idea of altruism overall (Hansen et al, 2021).

Other framings have been used so far that evocate a similar approach to altruism which are represented in the table below.

TABLE 7.

Comparison of regulatory incentives based on frequency.

Approach	Main Characteristics
Data *solidarity*	This approach should not be conflated with data sharing, as the main scope of the approach is to increase collective control, oversight, and ownership over digital data and resources. According to data solidarity principles, not sharing or not collecting data might, under certain conditions, be the best solution to protect certain groups and prevent harm (Prainsack et al, 2022).
Data *cooperatives*	These identify with the international cooperative movement and the bottom-up process to collect and manage data sets. Data cooperatives are organisational structures constituted by different data subjects. These entities help members of the cooperative to exercise their rights over their data (Fink, 2024; Micheli et al., 2023).

Approach	Main Characteristics
Data *collaboratives*	Usually established with a clear problem to be solved. Data collaboratives refer to a new form of collaboration, beyond the public-private model, in which participants from different sectors provide access to their data for (re)use in the public interest (Kalkar and González Alarcón 2023; Susha et al, 2018).
Data *philanthropy*	This approach describes the process of donating various types and forms of data by individuals and companies for the public good (Taddeo, 2016).

Source: own elaboration.

Although not aiming to be comprehensive, Table 7 shows some of the various forms of data governance and sharing models that have appeared in recent years. Whilst, on one hand, this is a clear sign of the need to bring innovation in data practices for accessing and reusing data that might be beneficial for the greater good; on the other hand, the table also highlights the difficulty to develop one single model that might respond to all needs for accessible and ethical data sharing practices and the need to see data sharing practice and intermediation more as a spectrum.

Although data altruism organisations have been defined as 'a new intermediary in the data value chain' (Baloup et al, 2021), it is important to state that data altruism organisations and data intermediaries have a clearly differentiated legal status and purpose under the DGA. However, their functions can sometimes be similar; as the table illustrates, many overlaps exist in practice and there is, in reality, a blurred line amongst all the different 'data mediators'.

The choice on behalf of the legislator to name the data sharing practice as 'data altruism' has raised few comments among experts and within the literature. By choosing to name existing organisations and data sharing practices as instances of 'data altruism', a concept not used before (or perhaps only marginally) has in fact raised critiques and started debates. Some experts (Lalova-Spinks et al, 2023) have questioned the actual novelty of the data altruism approach and considered whether building on already existing models and avoiding the creation of new forms of intermediaries (RDAOs) would have been a preferrable approach instead. Others (Prainsack et al, 2022) have criticised the choice of the concept of altruism altogether as a misnomer that assumes individual general self-interest as the norm, and altruistic acts as an exception. Moreover, the very notion of altruism seems to imply that self-interest and concern for others are mutually exclusive, whilst in real life it is often the case that both exist jointly.

The second key concept in the DGA definition is the notion of ***general interest***:

> *There is a strong potential for objectives of general interest in the use of data made available voluntarily by data subjects on the basis of their informed consent or, where it concerns non-personal data, made available by data holders. Such objectives would include healthcare, combating climate change, improving mobility, (...).* Art. 45 of the DGA

Unfortunately, the DGA only provides a non-exhaustive list[13] that – it has been argued – appears to create different levels of general interest (Finck and Mueller, 2023), namely:

13. Article 45 lists the following objectives as exemplar '(...) healthcare, combatting climate change, improving mobility, facilitating the development, production and dissemination of official statistics, improving the provision of public services, public policy-making, and scientific research'.

general interest linked to thematic area, on the one hand, and/or research and public interest purposes on the other. Although it seems clear that general climate change data would fit into the data altruism model, it remains unclear, for instance, whether climate data for mitigation and adaptation would fit into the model or what other positive climate objectives could be pursued (e.g. promotion of biodiversity, etc.) with the same data.

Moreover, notions of general interest are not easy to address as these are deeply political concepts, that require political choices and depend on political preferences (Baloup et al, 2021). What falls into the notion of 'general interest' (and what does not) – such as, for instance, different aspects of climate change and climate action – might vary in different countries and at different points in time, based on the political will. To avoid that these definitions of general interest end up being tautological and empty (as one can only tell that the general interest is what it is in the interest of the general public), a greater use of participatory methods and deliberative democracy approaches (OECD, 2020; Fishkin, 2009) could be recommended. Engaging the public (e.g. through mini-public methodologies) might contribute to identifying and building shared principles and definitions of what the public values are and what might be in the general interest. A noteworthy attempt in this direction is the example of the Scottish government, which in 2022 gathered the public opinion to co-design a series of principles for unlocking the value of Scotland's public sector data for the public benefit[14].

14. 'Unlocking the Value of Public Sector Data for Public Benefit' is a programme to unlock the value of Scotland's public sector personal data via its use with or by the private sector, for public benefit. The programme adopted a citizen-led approach to co-design principles for data sharing practices. More info: https://www.gov.scot/groups/unlocking-the-value-of-public-sector-data-for-public-benefit/

4.2 Methods

Starting from the overall objective and scope of the research as defined above, the research questions identified and addressed in this chapter are the following:

1. What is the state of the art of data altruism in the environmental data space?
2. Who are the key actors that could best support the data altruism approach to become a reality (with a special focus on environmental data)?
3. What are the possible power imbalances that might emerge and hinder data altruism approaches?
4. What is the potential and what are the challenges for a successful implementation of the data altruism model?

This chapter follows a qualitative research approach, which includes the following methods: (i) a literature review exploring the key topics and the current debates/discourses through a critical lens; (ii) desk research and document analysis, including policy briefs and grey literature; and (iii) a series of semi-structured interviews with key experts in the field[15]. As the process for the establishment and conditions for RDAOs are still unclear, the experts' engagement became a fundamental step to explore possible future use cases of data altruism within the GDDS, as well as to explore possible practical

15. These experts were selected from within academia and civil society organisations. Academics and civil society experts were selected for their work on the topic of data sharing and data governance. Others were selected as they could bring first-hand expertise of running data sharing organisations or initiatives or supporting data sharing mechanisms and organisations in different ways. The author of this chapter is truly grateful to all of them for sharing their time and thoughts with her.

future roles and activities of registered data altruism organisations for environmental data sharing, sketching profiles of who these organisations could be, what skills they will need, and what challenges they might face to gain the needed trust from citizens so that a data altruism model could thrive in the environmental field.

Moreover, whilst acknowledging the potential impact of the data altruism model for incentivising environmental data sharing from private companies, corporations will not be the main focus of this chapter. Exploring possible incentives and barriers for corporations to adopt a data altruism approach would need further research, which is outside of the scope of this work and currently limited given the novelty of the approach itself.

4.3 Data altruism in the context of the GDDS (state of the art)

Europe has an ambitious target of reducing its emissions by 55% by 2030[16]. But to achieve a just transition and meet the ambitious goals set by the Green Deal, more and better data and data access are needed to measure whether and how actors are making progress, to develop models for responding and (whenever possible) anticipating catastrophic events that happen as a result of climate change, and to improve research and knowledge that can establish a sustainable economy and advance climate justice.

Despite the intention to create a link between the European Green Deal and the EU Strategy for data (European Commission, 2020), this synergy seems to halt when it comes to substantive norms (Finck and Mueller, 2023). Digital data is central for the advancement of the Green Deal, as it provides key information for planning, measuring, and ensuring accountability on environmental measures. Environmental data is currently collected through multiple sources including satellites, production and supply chains, the IoT and connected devices that are used in private households, as well as other sensors and citizen science initiatives. But recent studies demonstrate that environmental data are currently mostly amassed by private actors as a valuable resource to build competitive advantages over other companies or towards public bodies (Verhulst, 2022b; Fritzenkötter et al, 2022).

The main focus and the novelty of the research presented in this chapter is the contribution towards a better understanding and future improvement of data altruism mechanisms for environmental data and for the purpose of advancing the GDDS. This topic has significant potential but is currently under-researched, as most evidence regarding the combination of the European Green Deal and the EU Strategy for data has focused on the use of digital technologies to achieve environmental objectives and only marginally on the use and value of data in this space (Finck and Mueller, 2023).

Finck and Mueller (2023) identify three main use cases for environmentally relevant data in the public sector and these pertain to activities that we could broadly define as: (i) anticipation and planning, (ii) delivery, and (iii) monitoring and evaluation.

16. See 'Delivering the European Green Deal' https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal/delivering-european-green-deal_en

- Destination Earth[17] or the project Future City Glasgow[18] are examples of environmental data used for *anticipatory* purposes of running simulations or building useful models to predict climate risks or disasters and support decision-makers in their planning role.
- Environmental data used for *delivery* would include the use of data to successfully implement environmental policies or other policies and interventions that might be relevant for environmental purposes. This is potentially a vast field, that could include mobility efficiency, energy reduction, biodiversity protection in agriculture, and many more areas.
- Data use in *monitoring and evaluation* activities would include the assessment of objectives, the identification of risks or unexpected effects of interventions, and policies that could have a significant impact on the environment.

Four characteristics can be identified that would be specific to data sharing practices within the environmental data space and in the fight against climate change.

Firstly, we need to acknowledge the **global nature** of both the problem as well as the solutions in the environmental space. Therefore, for environmentally relevant data sharing practice it would be critical to enhance the **cross-border data flows**, by ensuring harmonised practices within the EU and beyond. The DGA in Art. 21 already mentions the additional rules that would be required for safeguarding and use of data collected in third countries, and a strong EU leadership on issues of data governance and data sharing would be needed to trigger data innovation and stimulate cross-border data sharing for environmental data. Moreover, it would be interesting to explore whether there will be options for a transnational equivalent to the national RDAOs that would support and facilitate transnational data altruism practices. An interesting example of data sharing and collaboration on a regional level is the **Baltic Sea case study** (Kalkar and González Alarcón, 2023), which in 2015 saw the European Union government satellite data shared with private shipping companies in order to help them adapt to the changing winter conditions in the area (Sawyer et al., 2015 in Kalkar and González Alarcón, 2023). A similar approach could be replicated, enhancing cooperation at the regional level with the aim of advancing environmental data sharing and action to protect economic and social actors from new risks emerging as a result of climate change.

Secondly, data collection and data sharing of environmental data would arguably benefit from a **longitudinal approach**. An interesting example in this direction is the **Colombia Longitudinal Survey** (Kalkar and González Alarcón, 2023) (also known as ELCA in Spanish). Developed by the Universidad de los Andes, the initiative followed the same 10,000 Colombian households for 12 years with the aim of mapping and understanding social and economic changes. A similar approach would be possible to allow, for instance, the use of available household generated data to monitor environmentally relevant phenomena and/or behaviour over time.

Thirdly, the **nature and qualities of environmental data** are different from other

17. The Destination Earth initiative was launched at the end of March 2022 as part of the Green Deal data space activity. More information are available here https://digital-strategy.ec.europa.eu/en/policies/destination-earth

18. Future City Glasgow is an ambitious programme aimed at using technology to make life in Glasgow smarter, safer and more sustainable: https://www.glasgow.gov.uk/futurecities

data sets, as the amount of data is potentially huge, could be of a highly technical nature, and it would consist of a mix of personal and non-personal data as well as individual and collective data. Moreover, it has been noted that 'EU (data) law does not distinguish between data of environmental relevance and data without environmental relevance'. (Finck and Mueller, 2023), and without a clear definition it becomes more complex to understand what regulatory instruments and data sharing practices apply to what data.

Finally, as the definitions of what constitutes environmental data are continuously expanding it would be ideal to adopt an **open and flexible approach to the data altruism model in the Green Deal data space**. In practical terms, this might include taking an open definition approach for defining the purposes for which the environmental (and environmentally relevant) data can be shared through altruistic models, as well as imagining ways for continuously assessing the use of data for unanticipated purposes. In fact, new technologies for data collection and new types of data might emerge, that will pose new risks and opportunities and would require continuous assessment. Adopting an open (but safe) approach will be central to embrace a long-term perspective and provide a future-proofed space for data altruism for Green Deal data.

4.4 Main findings

The findings presented here draw on the literature, as well as the semi-structured interviews with experts. From these sources, whilst no strong voices have been raised against the data altruism model, many critiques have emerged around its successful implementation in the DGA context and regarding the tool of the EU and national registers.

While data altruism demonstrates significant potential to contribute towards the achievement of the Green Deal agenda – e.g. by facilitating the voluntary sharing of environmental and climate-related data – several risks can undermine its practice. From a technological perspective, aspects such as data security and data privacy concerns (e.g. deterring individuals from contributing due to fears around data misuse or identity theft), have been found critical. In addition, the value of data altruism may also be hampered by poor quality and integrity of the data, leading to inaccuracies in research and decision-making, and undermining the overall sustainability of the model. From a social and organisational perspective, incentives for participation are essential, yet oftentimes challenging to implement. Furthermore, navigating the complex legal landscape of data protection, especially across borders, can pose significant challenges, as do ethical issues related to informed consent and the rights of data subjects.

When examined through the lenses of the DGA, other issues emerge that risk undermining the data altruism model and its implementation across the EU, namely: (1) the fact that the model is new and not yet well-known; and (2) overall uncertainty regarding its implementation requirements across different EU (and non-EU) countries, including potential differences in the distribution of recognised data altruism organisations and support at national and EU level. These factors risk hindering both the supply and demand side of this potential new data sharing market, which might impact on the amount and diversity of organisations that could put themselves forward to take the role of RDAOs, and consequently on the amount and quality of environmental data collected and shared.

The first objection questions the need and built-in assumptions of the model overall, as scholars (Finck and Mueller, 2023; Veil, 2021) argue that equivalent data sharing practices and data altruism organisations already exist[19] that still remain in a grey area, as the DGA does not clarify whether and under what conditions certain data sharing practices are admissible and what organisations will be able to register. Although the development of a new EU umbrella concept (that of data altruism) is likely to benefit to data sharing practices and organisations in the long term, the current perception is that the DGA has mostly brought in new top-down requirements and procedures instead of clarity and support: 'If the EU had truly wanted to facilitate processing of personal data for altruistic purposes, it could have lifted the requirements of the GDPR, which are almost impossible for many controllers to meet'. (Veil, 2021, p4). With time it will have to be explored how broad and inclusive the concept of data altruism really can be and how easy it will be for existing organisations to align and recognise themselves in this model. The fact that – as the DGA establishes – the Rulebook itself will be developed with existing altruism organisations and in a collaborative manner is a positive sign towards a more inclusive framing of data altruism.

Baloup et al. (2021) present a detailed discussion about three further elements of the data altruism model which, they argue, remain dangerously undefined in the DGA: (i) the consent process, (ii) the application of the mechanism to personal and non-personal data, and (iii) the concept of 'general interest' and its difference (or not) with the concept of 'public interest' as expressed in the GDPR. Whilst the authors are positive about the potential of a planned European Data Altruism Consent Form to increase readability and understanding of the conditions of consent from data subjects, they argue that the DGA might bring additional uncertainty to the already unclear notions of purpose and processing activities that are involved in the giving of the consent (e.g. whether this is given for one or more specific purposes and/or for one or more data processing activities). Other scholars (Lalova-Spinks, Meszaros and Huys, 2023 and Baloup et al. 2021) even introduce the idea that a Data Altruism Consent could be considered as a whole new model of consent in its own, building on the new step-based approach of the Court of Justice of the European Union[20]. Alternative solutions point to the fact that adding an *altruism exemption* in the GDPR might have been a simpler and more effective solution (Finck and Mueller, 2023; Veil, 2021).

Most commentators and scholars that have been engaging with the concept of data altruism so far (Lalova-Spinks, Meszaros and Huys, 2023; Finck and Mueller, 2023; Baloup et al, 2021; Veil, 2021) seem to reach the conclusion that the data altruism model under the DGA might suffer from existing uncertainties (including data protection and

19. Examples would include: the Robert Koch Institute' 'Data Donation App' (https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Corona-Datenspende-allgemein.html), which allows data subjects to donate data to support research on COVID-19 or the Decode project (https://decodeproject.eu/index.html), which enabled citizens in Barcelona and Amsterdam to collect environmental data and make this available to the general public (cfr Finck and Mueller, 2023; Veil, 2021).

20. For more details see the CiTiP White Paper on the Data Governance Act (Baloup et al. 2021), which assesses possible interpretations of the DGA aim regarding consent practices, including the 'step-based' approach recently contemplated by the Court of Justice of the European Union (CJEU) in its recent case law on joint controllership. This approach seems to put forward a processing activity- (or operation-)based understanding of consent.

the interplay with the GDPR). The fact that the details of the implementation of the data altruism model and their registers are left to the discretion of national authorities (with different incentives and interests) also risks exacerbating issues of uncertainty in the operationalisation of data altruism.

Finally, the main assumption from the DGA involving trustworthiness as the most important element for enhancing data sharing practices is not fully developed and more research would be needed to better understand how trust develops differently among the different data subjects and how trust mechanisms work within and among different EU Member (and potentially non-members) States. Further exploration of the question of trust and what might undermine or reinforce trust on the part of different actors in different systems would be necessary. This would require more qualitative research that could explore actors' different incentives and motivations to take part and collaborate in the advancement of the data altruism model. Studies that explore qualitatively the attitude of key actors regarding the concept of data altruism are still limited. Lalova-Spinks et al. (2023) is in fact one of the few qualitative studies that engaged experts (rather than the general public or corporations) to map their attitudes regarding data altruism in healthcare, although its findings are potentially relevant to the field of environmental data. From the study, it emerged that actors are overall supportive of the idea of reusing health data, as long as a number of criteria are met: first, trustworthiness of those who will access, store and use the data; second, the perceived sensitivity of the data itself; and finally, the degree to which the data are expected to contribute to the public good. Skovgaard et al. (2019) interestingly show how concerns among people living in the European Union regarding data sharing in healthcare increase when the possibility for commercialisation of data is involved, which in turn raises issues of data security and of the potential use of data against the interests of the people providing the data. When transposing these findings to the environmental space, few additional considerations are needed. Firstly, as a significant amount of environmental data is held by corporations, further research would be needed to understand their motivations and possible incentives to share their data in altruistic ways. Secondly, the impact of data sensitivity in the case of environmental data would need to be explored, and other concepts (such as reputation and commercial interests) that could affect the possibilities for data sharing might need to be considered. Finally, differently from healthcare, environmental data might raise more political issues, for instance on whether and which modes of use and reuse of environmental data will be considered to be in the public interest, who might benefit or be harmed by the use of environmental data, and what evidence and sources of knowledge will be considered to be of more value.

4.4.1 Actors' framing and motivations in the GDDS

The mapping of the potential actors involved in the future environmental data altruism ecosystem should be understood extensively. As the data altruism model is still new and the ambition of collecting and reusing environmental data is huge, no potentially interested party should be excluded. Future use cases and scenarios will have to be developed at the EU level (and transnationally) to expand our imagination of the possible roles of RDAOs and interactions among and with other key actors. A general list might include:

- **The sponsors:** those who will be responsible to promote and raise awareness of data altruism in the environmental space;
- **The collectors:** those who will be responsible to collect and store the data safely;
- **The data sharers:** those actors (e.g. individuals, businesses, academics, and other organisations) who need to be aware and motivated to share their data;
- **The data users:** those who will ensure that the right data is used in ways that are most beneficial for environmental purposes.

The following pages will further explore possible roles, challenges and opportunities for data altruism sponsors and collectors, which are potentially new actors that will play a key role in the initial phases of the dissemination of the data altruism model and the uptake of the EU and national registers.

Sponsors of the data altruism model would obviously include the **EU institutions as well as national and local authorities and governments** who would have a key part to play to ensure resources, infrastructure, and political support are given to ensure the growth of the data altruism model also through the publication the EU and national registers. Another key (but perhaps less obvious) potential actor that could play a role in the data altruism ecosystem (as a sponsor as well as data sharers or users) are **environmental social movements**. These should be understood as a wide constellation of different actors, which will include on a spectrum more activist and agonistic groups, as well as informal groups that campaign for healthy food consumption or consumer movements. Social movements, which have so far been the most successful to ensure that climate change is at the top of the political agenda, could play a role in increasing data sharing practices for environmental purposes. First, they would ensure the inclusion of diverse and alternative knowledge (della Porta & Pavan, 2017) in a domain – specifically, climate change – that is still highly contested and polarised, especially in certain countries. The deliberate engagement of social movements in the dissemination of the data altruism model for environmental data could also ensure a higher standard of social justice and equality (Parks, 2020); provide alternative spaces for debating the issues of environmental data sharing and use in political terms (Corry and Reiner, 2021); as well as continue pushing for ensuring that climate change stays at the top of the political agenda in different countries (Walgrave and Vliegenthart, 2019).

The collectors of the data altruism model would of course include the RDAOs, which is the main novelty introduced by the DGA and which has high potential to influence the success of the data altruism model overall. They are intended to play a key role as the designated intermediaries for the data altruism space but are, at the moment, still undefined. Intermediaries are generally defined as third-party actors who help match supply and demand between data providers and data users. Successful data sharing and (re)use should consider all the phases of the data lifecycle, which would include the planning, as well as the storage, analysis, and reuse of data (Kalkar and González Alarcón 2023). In this respect, it has yet to be clarified whether and how RADOs would play any role in the planning phase, for instance by analysing and giving shape to the demands of environmental data for specific purposes. Shaping the data demand and planning data collection in an intentional way is not an easy task. It would require these organisations to identify current and future environmental data needs, as well as having

an overview of what data are available and the possible nature of the data (and whether or not such data can be used in the 'general interest').

Experts also pointed to the potential problems with the representativeness of data collected through data sharing practices, as this might not satisfy the diversity of data and data sources that would be needed. Reaching out to diverse demographic groups and developing the right technologies for facilitating data sharing (especially for environmental data) would be a necessary measure to ensure that a better sample and a more useful set of data is collected. Both aspects require expertise, time, and resources that RDAOs might not have at their disposal.

Even if the role of these collectors is still undefined, a few considerations can be put forward based on the literature and informal discussions with experts regarding challenges and opportunities to become RDAOs that should be addressed by the EU Commission in the next steps.

4.4.1.1 Resources and skills

Ensuring enough resources and skills will be the main challenge for RDAOs. If not properly supported (through funding, training or other means), these organisations risk, in fact, to fail.

Building on previous examples of data sharing platforms and processes (such as *tracking. exposed* and *AIForensic*), experts also raised concerns about the fact that one single organisation could concentrate the right skills, technologies, and resources to do data collection, anonymisation, and analysis at a high-quality level. As experts suggest, these three actions could be better implemented by specialised individual organisations, which together could form the infrastructure for a data altruism model to thrive.

4.4.1.2 Develop viable business models for RDAOs

The question of how RDAOs can gather enough resources to maintain themselves is central (TEHDAS, 2023), as these organisations will need to hire the required expertise, build new and maintain old technologies, as well as do the networking, educational, and promotional work that is needed for their mission to be successful. One role for those actors who will be sponsoring and supporting the data altruism ecosystem will be to identify viable and socially driven business models and provide examples and support for RDAOs to find the right model for them and to apply it successfully. The Digital Impact Alliance[21] is an interesting example of a multi actor **initiative to raise awareness amongst major donors** regarding the need to invest in the data sharing space. Similar initiatives will be needed at the EU level to foster funding opportunities for data altruism and for successful RDAOs.

4.4.1.3 Terms of services

Developing detailed terms of services for RDAOs would be an important step to ensure quality and consistency across different RDAOs at the national and EU levels. These terms of services should be established in collaboration with all actors involved and through the development of use cases. These will allow to identify currently unanswered questions, like for instance: (i) how and when the opportunity

21. The Digital Impact Alliance recently launched a Climate Data Joint Learning Network, which has the following aim 'By bringing together insights from the emerging field of digital public infrastructure (DPI) with experts on climate action, the JLN will surface unique recommendations for funders and other actors who can accelerate the use of data to meet the urgent need to build community resilience in the face of climate change'. Accessed on 12 June 2024 from their website: https://dial.global/work/joint-learning-network-unlocking-data-for-climate-action/

might emerge to prompt private actors to share their data in the first place; or (ii) what might happen to the collected datasets if a RDAO ceases to operate. A detailed journey map through all the possible stages of data sharing via data altruism organisations might be a useful reference for RDAOs in the future.

One model that has proven successful in gathering data is the **open innovation challenge approach**, where actors are prompted to share their data within a competition-based model to respond to a public challenge and provide innovative solutions to public problems. Although the competitive element of this approach might not always fit well with an 'altruism'-based model, it might still be worth exploring. Environmental challenges in fact could be a beneficial tool to prompt a demand for data sharing in specific fields through open calls. This would also equip RDAOs to play a more intentional role in the planning of data collection.

4.4.1.4 Specialised vs generalist

Contrary to data intermediaries which need to remain neutral, RDAOs can take a stance on the data that they want to collect, the purpose for which they collect them, and the use they will promote. For instance, RDAOs will have to consider the option of being generalists, and collect all data that are shared with them, or position themselves as specialists, operating in one specific domain, in specific use cases, type of data, or even potentially in their geographical focus. They could make these decisions based on their mission (what we can a **Mission First Model**) or be more opportunistic and look first at the data that are available or prioritise the data sharing as a principle and achievement in itself and independently from the use (**Data First Model**).

4.4.2 Insights on power and trust

The data altruism approach should be considered as one way to address and challenge the data power that is now concentrated in the hands of few actors. Data (and especially environmental data) are in fact still mostly locked away in corporate and privately owned databases (Finck and Mueller, 2023; Kalkar and González Alarcón 2023; Fritzenkötter et al, 2022), and still unavailable and therefore not usable for purposes of general interest.

Questions of power in data use and sharing include (among others) the issue of asymmetries that an unequal data distribution, access and knowledge could enhance. Verhulst (2022) identifies two different types of asymmetries: from *data asymmetries*, where those who might need the data do not have access to them, to *agency asymmetries* that refer to imbalances in data collection and (re) use relationships, where vulnerable groups might be becoming further disenfranchised by not having full and informed access to data sharing processes. One way to address these asymmetries and acknowledge the power issues that are implied in data altruistic models is to ensure that justice principles are embedded in data sharing mechanisms 'by design'.

Kalkar and González Alarcón (2023) put forward a simple suggestion to include **the principle of 'targeted transparency'**[22] to inform the design of data sharing mechanisms. Other practical suggestions

22. Taken from the original text *targeted transparency becomes an important approach to promote data flows and data sharing under scenarios where the traditional mechanism of transparency fails (...).* (p.9). As clarified in the original text, targeted transparency refers to publicly required disclosure of specific information in a specific format in order to achieve a clear public policy purpose.

would include adopting a design justice approach (Costanza Chock, 2020) that would require the participation of those who might be affected in the design of the data sharing mechanisms and in deciding the principles that will inform the data altruism approach. A more participatory definition of inclusive and just design principles – similar to those identified by the Scottish government (see footnote 3) – although not a panacea, would go in the right direction to ensure certain values are upheld, towards a more data just approach for data sharing.

Other scholars (Lalova-Spinks, Meszaros and Huys, 2023) have even argued that the framing of data altruism might put too much responsibility on citizens overall, as they should choose how and when to voluntarily share their data without the right information and support. Increased citizen information and critical digital literacy (Ragnedda, 2018) might therefore be preconditions for a data altruism model to succeed. The risk is otherwise to implement a model of data sharing that only empowers those that already are (Gurstein, 2011). **Critical data literacy and clear regulatory boundaries** will ensure that citizens have the information, the confidence, and the opportunity to take part in the data sharing process, as well as opting-out of these processes when there is not enough trust. Whilst Art. 52 of the DGA usefully mentions the need for user-friendliness in the process of granting as well *withdrawing* consent to altruism organisations, the Article does not mention the possibility of *non-consent*, which might instead contribute to increased protection of data subjects as well as increased trust (through the recognition of the importance of *not trusting*).

Issues of trust are critical to the success of data altruism models. Whilst the EU perspective on 'altruism' seems to be underpinned by the idea that a 'certified trustworthiness' will create sufficient incentives for data sharing (Prainsack et al, 2022), what trust means for different actors and how trust can be gained and maintained should be understood more holistically. As one expert commented in the interview, those people who are 'altruistic' enough to want to make 'their' data available to a good cause might not need any further incentives. If the gold standard of the GDPR does not create enough trust in the processing of personal data, it is questionable what role the registration would play in the trust-building process. According to this view, the registration mechanism alone might not be enough to build the trust that is needed to motivate more people or companies to share their data voluntarily. What other and different incentives might be needed to achieve the purpose of building more trust is a key question that would have to be further explored. Also emerging from the expert interviews was that the question of trust is usually mentioned with regard to the people who should have trust in the process; but what is not mentioned is how and why we should have trust in the data. With this in mind, citizens and the general public should be encouraged to become more critical of the quality of the data being shared (whether what is shared is genuine and properly collected and presented) and perhaps even question whether data are trustworthy as a (sole) source of knowledge. These are also important questions that would need further exploration in the future steps of the data altruism model.

4.5 Conclusions: lessons learned and recommendations

A point that clearly emerged from the literature and interviews was that a prerequisite for successfully establishing the data altruism model revolves around the need of reducing the uncertainty surrounding RDAOs and their establishment. This aspect involves further clarification concerning the mechanisms for sharing data with organisations, how exactly the collected data may be used, and the legal framework and the compliance costs resulting from it: 'The related uncertainty means that potential creators of RDAOs as well as potential providers and users of the related data will think twice about whether they want to incur related risks and costs'. (Finck and Mueller, 2023).

The publication of the Rulebook for Data Altruism Organisations will be instrumental for providing further clarifications on the role and conditions for the RDAOs[23]. In addition, future considerations from the Commission regarding the **provision for adequate resources** to cover the costs associated with collecting, processing, managing, storing, and using data in safe and effective ways would be highly beneficial. These costs, as well as the need for infrastructures, expertise, and skills that the RDAOs will have to use should not be underestimated, as well as addressing the concerns that competitors might exploit the data collected by RDAOs for different and inappropriate ends (Finck and Mueller, 2023).

The DGA's label alone might not be enough to **incentivise the voluntarily sharing** of data from private actors (Finck and Mueller, 2023). In the face of many potential barriers and uncertainty, the lack of incentives to create RDAOs could be another key factor limiting the adoption of the data altruism model overall, and for environmental purposes specifically (Finck and Mueller, 2023; Veil, 2021). Incentives will have to be explored for the different data subjects (including companies) as well as for data sharing organisations.

The **development of *use cases*** will also be crucial in the process of operationalising data altruism as a way to help refine concepts and prove them in real use and with real actors. Whilst some examples of data altruism in prac¬tice have been listed (Veil, 2021), including, for instance, the DECODE project (see note 6), or ***Open SCHUFA***[24] and the *Corona Data Donation App*[25], whether and under what conditions these initiatives will be permissible and can be successful within a RDAO model for environmental data is still unclear.

Based on the findings and conclusions, this chapter highlights four key recommendations towards the implementation of the data altruism model:

1. Find practical ways to make the data altruism more accessible;
2. Shift the thinking from list of actors to a data altruism ecosystem;

23. Finck and Mueller (2023) also warn of the risk that rather than reducing compliance costs, the DGA's regime on RDAOs might increase these costs, for instance, through additional reporting requirements on the organisations.

24. This project by Algorithmwatch and the Open Knowledge Foundation examined the SCHUFA, Germany's largest credit enquiry agency. More than 4,000 people donated their SCHUFA self-disclosures in order to enable a check of the score determined by SCHUFA through reverse engineering.

25. This app from the Rob¬ert Koch Institute (RKI) was downloaded by 530,000 persons in Germany (as of 30 September 2021). Participants linked their fitness wristband or smartwatch and gave their consent to scientific data analysis. Measured values such as sleep patterns, heart rate, and num¬ber of steps were then transmitted to the RKI, help¬ing scientists to better understand the spread of the coronavirus.

3. Enhance public engagement and oversight;
4. Ensure collaboration vs competition in data altruism.

4.5.1 Making data altruism accessible

Data altruism success will rely first and foremost on legal and non-legal persons and entities to voluntarily engage in altruistic data sharing of environmental data and, most importantly, to do so by providing data that are accessible, useful, understandable, and processable by RDAOs.

As suggested elsewhere (Kalkar and González Alarcón, 2023), the development of *data stewardship capabilities* would be relevant in a data altruism model in order to better identify opportunities for data reuse and for conditions of general interest. *Data stewards* might be a needed support raising awareness, as well to navigate the complexities of data altruism for environmental data and for steering responsible data sharing and data reuse within data altruism initiatives. Kalkar and González Alarcón (2023) go one step further in suggesting the use of the Contractual Wheel of Data Collaboration as a key tool for data stewardship conversation. The wheel provides a simple but comprehensive guidance to walk actors through discussions across six guiding aspects of data sharing decisions (the why, who, what, how, when and where) to explore: why data need to be shared, what data and what formats are used, who is involved and what are their responsibilities, how data and the relationships with data subjects are managed.

4.5.2 From actors to ecosystems

Considerations from a system viewpoint on data altruism implementation would include thinking beyond the key actors to include questions of governance and incentives, as well as processes and long-term resources. Already criticised for its top-down and bureaucratic approach (Veil, 2021), the RDAOs model risks failing if unnecessary bureaucracy is put in place that will raise the bar and the running costs and might disincentivise interested organisations from applying to become RDAOs. A focus on ways to build a strong data altruism ecosystem might be more efficient in ensuring the right level of protection of data, as well as reducing unnecessary bureaucracy. In practical terms, an ecosystem approach would consider ways of balancing different interests at stake from all stakeholders involved, including balancing the need for bureaucracy and new procedures with the need to build infrastructures and financial incentives for RDAOs. Practical proposals have been put forward (Veil, 2021) that include introducing an exemption for facilitating the processing of data for altruistic purposes, or the idea of establishing an 'Altruistic Controller' that would take into account the public benefit purpose for data processing. A long-term view and an ecosystem approach might also help address the issues of *data legacy* for ensuring that the environmental data collected would be available and reusable in case the designated RDAO who initially gathered them might cease its activities for various reasons (e.g. including lack of resources).

4.5.3 Enhance public engagement

Public engagement and oversights would contribute to address the issue of trust, as well as making the altruistic model of data sharing more democratic and inclusive. The call for public engagement in 'datafied' societies is on the rise (Warne et al., 2021), and several methods have been developed and used to

engage citizens in deliberations on key topics and for the purpose of advancing democratic innovation. Not only would public engagement in the designing of data sharing innovation be important for democratic reasons, but citizens' active engagement should be also sought in monitoring these models to improve trust and acceptance of the principles of data altruism as well as of the role of the RDAO over time. When it comes to environmental data, this citizen-centred approach would also be in line with the Sensing for Justice (SENSJUS)[26] project to advance grassroots-driven environmental monitoring as a source of evidence (Berti, 2024).

Providing innovative modes for public engagement and information to data subjects and the public at large on how collected environmental data has been used in research and the impact it has achieved would also be recommended (Hansen et al, 2021).

4.5.4 Ensure collaboration vs competition

As highlighted by many experts in the informal interviews, there is a risk that RDAOs, if not properly implemented, could create a system where competition for data sharing and other resources shapes the data altruism model. The proliferation of too many and competitive RDAOs would in fact result in further fragmentation of environmental data, which would be detrimental to the overall aim of advancing climate action. Designing modes and incentives for collaboration from the start (also building on already existing models of data sharing) would be essential for avoiding unintended negative effects of data altruism on environmental data. Among others, approaches of data solidarity and data collaboratives could be considered, including providing an avenue for increasing the collective bargaining power of individuals in ensuring data sharing practices according to certain agreed values (e.g. cooperation).

As stated in the introduction, whilst waiting for the Rulebook and EU and national registers to be in place, the aim of this chapter was not to provide a full analysis of the impact of RDAOs on EU data sharing practices but to provide an initial and introductory examination of the potentials and challenges of the data altruism model in general, and specifically to sketch a future practical role and use cases for RDAOs. Further research would be needed in the future to explore the effective impact of the data altruism model in the environmental space and on the role and incentives for corporations as key actors to unlock the true potential of environmental data sharing practices.


26. The Marie Skłodowska-Curie Actions and formerly the Dutch Research Council funded the 'Sensing for Justice' project, hosted by the JRC within the INNPULSE project from 2020 to 2023. https://cordis.europa.eu/project/id/891513

5 BUSINESS DATA SHARING AND THE GDDS

5.1 Introduction

The European strategy for data aims at establishing a single market for data in the EU. To achieve this, it envisages a common data space in the European Union, made up of sectorial data spaces whereby data is shared, pooled and flows among different actors, especially those in the private and public sectors. Among these is the Green Deal Data Space (GDDS). The GDDS also aims to implement the EU's European Green Deal, to utilise data to achieve its objective of Europe becoming the first climate neutral continent, and tackle climate change, in particular by supporting the data economy as a means of achieving Green Deal objectives in zero pollution, a circular economy, biodiversity, and climate change/climate adaption.

Underpinning these objectives are the assumptions that: (a) the private sector will be involved in data spaces; and (b) that the access and use of (pre-existing) data can be achieved and supported through business models and can also in turn lead to positive outcomes including as regards climate change. Both of these assumptions must be interrogated, as they do not necessarily and logically follow from the establishment of data spaces nor the increased access, sharing, and use of data. In fact, as regards the latter, there is significant literature and evidence pointing to the costs of digital data, its use and its associated innovations such as artificial intelligence (AI) for the climate (e.g. Brevini, 2021; Lopez Solano et al, 2022), which must be mitigated if data itself is to be considered climate neutral. This has been acknowledged by the EU, and the European Green Digital Coalition has produced a methodology in order to ensure ICT solutions are employed to deliver net positive carbon impacts (European Green Deal Coalition, 2024). Nonetheless, some benefits from ICT developments, such as AI, may be speculative or difficult to quantify accurately at this point in time (see Luers et al, 2024), and so these limitations must be taken into account. Indeed, the encounter between data and sustainability is multifaceted, as recognised by Noto La Diega and Derclaye (2023, p.6) who define 'data sustainability' in three ways: as the use of data to monitor progress vis-à-vis the UN's Sustainable Development Goals; the use of data and digital technologies to achieve more sustainable outcomes; and the process of 'overcoming unsustainable data practices' (p.6) – all of which should be accounted for in the GDDS.

This contribution focuses on the first assumption, that businesses will (want to) be involved in data spaces, by considering the understudied area of business data sharing. In particular, it aims to **answer the research question of whether and what opportunities and challenges exist for private sector actors to participate in such data spaces, with a focus on the GDDS**. It also offers a critical appraisal of the role of the private sector in data sharing and whether and in what form it is desirable to achieve data and environmental justice outcomes.

To accomplish this, socio-legal/law in context methods are used, drawing on literature, policy documents and other contextual material and analysis from disciplines including law, policy, critical political economy, and critical data studies. The method principally involves desk-based research. This research has taken place between December 2023 and June 2024. From here, some background on this issue will be provided, starting first with existing 'green data', and B2G, B2B and B2C data flows, which may inform business involvement in the GDDS. Next, opportunities for businesses to

serve the public good through corporate social responsibility and environmental reporting are considered, before moving to challenges that principally come from legal regimes that may block such data sharing, such as intellectual property, confidentiality, and competition law. From a socio-legal analysis of the current situation and materials, overall the urgency of climate change and business corporate social responsibility may be persuasive to businesses to facilitate more sharing of their data, if this is done against a backdrop of clear rules from the outset governing the GDDS and addressing points of uncertainty around the application of other areas of law such as competition, and liability issues. If this, however, does not lead to the desired level of data sharing on a voluntary basis from business, the EU should consider new obligations which would mandate the private sector to share data that is required for the GDDS to achieve its climate change objectives.

5.2 Context

In this section some background and context are offered on business data sharing and the GDDS, as regards what kinds of data may be in the GDDS, current legal and policy frameworks for environmental data sharing, the private sector organisations which may be involved, and existing research on data sharing, encompassing B2G, B2B and B2C data flows as all of these may take place within data spaces (along with the flow of public sector data to these groups as well, although that is beyond the scope of this report).

5.2.1 Green data

Data relevant to the GDDS held by the private sector can take several forms. They could be personal and non-personal data, although 'there is a higher degree of non-personal data' in the agriculture and energy sectors, which may be (more) relevant to the GDDS (Comandé and Schneider, 2022, p.742). There may also be mixed datasets which contain both personal and non-personal data.

Geospatial data is a paradigm example of environmental information or data, and forms the basis of the INSPIRE Directive, the objective of which is to create an infrastructure for spatial information in the EU, using government data for EU environmental policies and activities[27]. These data are envisaged as forming an important part or source of data for the GDDS in the form of the public sector contribution, but there are limitations: to this sector, and to a particular kind of environmental data.

Other public sector data may be available under the **Open Data Directive**, which has as an objective the creation of common rules for an EU market for public sector data by making such data available and reusable[28]. The Directive recognises a particular subset of public datasets called 'high-value datasets', which include various categories of environmental data including geospatial, earth observation and environment, and meteorological data. Other categories of high-value datasets that may include environmental data are statistics and mobility.

It is important to note that the distinction between public and private sector data is not always clear in practice, and the distinction is not the same thing as 'public domain vs proprietary data' as public

27. Consolidated text: Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE) (OJ L 108 25.4.2007, p.1).

28. Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the reuse of public sector information (recast). PE/28/2019/REV/1. OJ L 172, 26.6.2019, p.56–83.

sector data can also be proprietary (OECD, 2019). Espinosa Apraez (2021, pp.2-3) also questions the hitherto 'binary' personal/non-personal and public/private distinctions made by the EU, as this 'does not always reflect the current dynamics of data production and can lead to counterproductive outcomes' such as vis-à-vis the private sector producing data 'of public relevance'. Notwithstanding these complexities in practice, the research considered current data sharing dynamics in regulation and policy for B2G, B2B and B2C flows, which may be relevant to green data and/or the GDDS.

5.2.2 B2G

Understanding who the private sector or businesses are, which are in turn envisaged to participate in the GDDS, is important to understanding the kinds of data they already possess and what they may contribute to the data space. There may be companies which already use environmental data such as the aforementioned high-value datasets and therefore may be well used to engagement with the public sector providers of such data. Other non-sector-specific businesses may include transnational 'Big Tech' and digital infrastructure companies such as Google and Amazon, which possess extremely large amounts of data on a myriad of topics, including 'green' matters. As Alemanno (2018, p.185) argues, 'private data remains the prerogative of a few big corporations who jealously guard it'. There are further potential participants in specific subsectors of green data, notably with interests in the circular economy, biodiversity and zero pollution. For example, in order to comply with the new EU Regulation 2023/1115 on deforestation-free products, traders in, e.g. coffee will need to engage in due diligence to show that their product, grown in another part of the world by contractors or sub-contractors at another rung of what can be a long and complex supply chain, does not contribute to deforestation and will need data in order to evidence that, which may be gained from the GDDS. Conversely, if such data are being gathered for compliance with due diligence requirements, it could form a source of data for secondary uses in the GDDS.

In any case, the data sharing dynamics and participants in the GDDS are likely to build upon existing examples and frameworks for business sharing of data. There have been several relevant policies and new laws developed in the EU in recent years. The European Commission in 2018 published a set of principles on private sector data sharing in B2G contexts (European Commission, 2018), which were then revised by the B2G Expert Group, including new principles on accountability, and fair and ethical data use, which should form the 'backbone' of B2G data sharing (High-Level Expert Group on Business-to-Government Data Sharing, 2020). **More recently, the Data Act, which came into force in early 2024, contains provisions (Arts. 14 and 15) to secure public sector access to data held by the private sector (and other non-public bodies) in cases of 'exceptional need'**, which comprises both public emergencies (including 'emergencies resulting from natural disasters including those aggravated by climate change and environmental degradation' according to Recital 64) and certain non-emergency situations (European Union, 2023).

The gains for the public sector from access to business data are elucidated by Switzer and Berti Suman's (2023) citation of the Analytical Report on Business-to-Government Data Sharing:

> The Analytical Report; Business-to-Government Data Sharing' (2020) prepared for the European Data Portal posits (again with a more prominent B2G focus) that *re-using relevant privately held data increases the public sector's ability to understand, assess and predict different situations and phenomena that affect the citizens. It enables more logical and fact-based decisions, and at a higher pace. The benefits from data reuse are not only reserved to the private sector. In fact, to become more cost-efficient and to provide effective services for citizens, public sector bodies can benefit greatly from data sharing and need to exploit the potential of new data sources.* The document also provides a relevant discussion of models and examples from real world case studies.

These case studies in the Analytical Report include Uber's Movement platform, which was a 'web portal providing data from over two billion trips … Anonymised, aggregated data [was] made available to the public as open data' (European Data Portal, 2020, pp.9-10). According to the Report, Uber launched Movement to 'attenuate' the 'tension' between itself, traditional taxi services and the public sector; it then closed Movement in 2023. The document sets out five B2G data sharing models: multi-party data sharing agreements; data donorship; data partnerships; data intermediaries; and data sharing by regulation. Some examples are given for data donorship including the aforementioned Uber Movement and, also relevant to the GDDS, Twitter (now X) providing UN Global Pulse with access to their data tools to support efforts to achieve the UN SDGs.

There has been little empirical research to date on voluntary B2G data sharing, sometimes known as 'data collaboratives' when they are aimed at solving societal challenges (Susha, Janssen and Verhulst, 2017). Some voluntary B2G data sharing has occurred in the context of crises and emergencies (see Holton, 2018; Susha, 2020). Indeed, the emergency scenarios necessitating B2G data flows are recognised in the aforementioned Data Act provisions. Susha (2020, pp.10-11) also identifies a list of critical factors for data collaboratives from a literature review:

Organisational factors

1. Appropriate and cost-effective business model
2. Articulating a clear and compelling value propositions to stakeholders
3. Availability of financial (and human) resources
4. Strong consortium with all required capacities and partner complementarity and fit
5. Alignment of incentives of the participants
6. Shared understanding of objectives, values, and expected outcomes
7. Matching the problem with the data or data insights needed to address it
8. Long-term strategy (data sharing strategy)
9. Clear measurable outcomes
10. Structured approach with clear (but flexible) agreements and regulatory mechanisms

11. Broad participation of all affected stakeholders throughout the process

12. Top management support

13. Building trust and investing in the relationship

14. Facilitative leadership (via a formally assigned manager)

15. Clear definition of responsibilities and process steps (iterative process)

16. Continuous mutual adjustment of partners to each other and adaptation of their roles (interdependence)

17. Open and regular communications (personal contact)

18. Commitment of stakeholders to the process

19. Evaluations (to identify 'small wins')

20. Contingency planning

21. Adequate technical/analytical skillsets and multidisciplinary teams

22. Fast delivery of results

Technological factors

23. Compatibility of technical infrastructure and interoperable standards

24. Using a systematic and transparent process to data sharing

25. Using simple and familiar data sharing infrastructure

26. Common concepts and terminology to enable data integration

27. High quality of data

28. Innovative analysis tools

Legislation and policy factors

29. Adhering to standards and community norms, including privacy and security

30. Appropriate risk sharing and effective risk mitigation strategies

Source: *Susha, 2020.*

In Susha's empirical work, she found that 'data quality and incentives have the greatest combined influence over the success of a data collaborative', and that shared understanding, value proposition, and trust were also of high importance (Susha, 2020, p.14).

Outside the emergency or crisis scenario, some further voluntary B2G examples can be found in Klein and Verhurst (2017), with a focus on official statistics. They point to the following incentives for private companies to share their data voluntarily with the public sector: 'mutual benefits accrued from working with National Statistical Offices (NSOs), the potential to develop new analytical skills, improve their reputations, generate revenue, meet regulatory compliance and demonstrate corporate responsibility' (Klein and Verhurst 2017, p.8). Such benefits for businesses are echoed in the Analytical Report for the European Commission's Data Portal (2020), which adds 'reciprocity' as another benefit, i.e. that private sector organisations through B2G data sharing can 'reciprocate [...] for the value they get' from the public sector's

provision of 'physical and digital assets' and public services, and recognises the situations where there may be a 'solid and sustainable' business case for the public sector paying for access to private sector data (pp.9-10).

Van der Sleen (2022, p.3), in a case study of the Dutch national statistics agency's data sharing partnerships with the private sector, found 'linked interests' between the public and private sectors when 'the self-interests of organisations are connected to create social value'. If there is a 'sense of urgency to share data', then this can incentivise the private sector to do so (such as in the early years of the ongoing COVID-19 pandemic) and can form a 'common ground on which the partnership can build' (Van der Sleen, 2022, p.3). The 'identification of a partner's needs, goals, and perception of the partnership have been identified … as [this] allows government organisations to better anticipate common interests and benefits'. Finally, 'establishing and maintaining intensive contact with partners'. Van der Sleen (2022, p.3) considers that all these elements contribute to creating 'mutual trust', an important element in voluntary B2G partnerships, which can flow from the data laws to which government organisations are specifically subject including vis-à-vis privacy and confidentiality of data. Van der Sleen (2022, p.3) also found 'the provisions of incentives and rewards' to 'create mutual benefits and a more sustainable partnership' which he found to be: 'data reciprocity (e.g. returning data to the data holder), knowledge sharing (e.g. on data processing and analysing methods), and 'cost reductions for data holders (e.g. quid pro quo agreement)'. Van der Sleen (2022, p.4) suggests governments focus on the urgency of data sharing and points to 'sustainability topics and resulting societal challenges' as a topic of increasing urgency which may incentivise voluntary B2G data sharing partnerships, and clearly is relevant to the GDDS.

Indeed, **sustainability goals have already given rise to voluntary B2G data sharing**. Susha and Gil-Garcia (2019) discuss Data for Climate Action (D4CA), a data collaborative which emerged from the Paris Agreement as a UN Global Pulse initiative in 2017. D4CA combined private sector actors with researchers from academia and non-profits in a challenge to fight climate change using data through the companies voluntarily sharing their data with the researchers. The winning team tackled air pollution in Mexico City using traffic data from Waze, and had both researchers and public policy officials as team members.

However, there can be differing incentives or disincentives on the public sector side. The public sector is diverse, and in the EU context can encompass the EU institutions themselves, national government-level institutions, regional or sub-national-level state institutions, and local authorities and municipalities, of differing sizes and with different powers. Smaller organisations in the public sector may be less attractive for data collaboratives with the private sector. Indeed, in a study of European municipalities' experiences with B2G data sharing, Micheli (2022, p.6) found disincentives in the form of the size of the public sector organisation: 'All respondents stated that accessing private sector data was challenging because companies often have no interest in selling data or in sharing it with a municipality'.

This reluctance to share data voluntarily is echoed by Mercille (2021, p.5):

> *In short, companies are reluctant to share data, which they see as a corporate asset.*

> *Additionally, when they do so, it is often based on self-interest, and may thus be unreliable to meet public needs. Therefore, emphasising the benefits of B2G data sharing to encourage companies to share data may be less productive than expected.*

B2G data sharing is provided for in the new suite of EU legislation on data, both compulsory and voluntary. As mentioned above, the Data Act includes provisions on mandatory B2G data sharing in certain circumstances of exceptional need, which Noto La Diega and Derclaye (2023, p.27) argue 'should encompass prevention of climate catastrophe' (p.27). There are other examples of pre-existing mandatory B2G data sharing, mainly concerning regulatory reporting, including for environmental matters (see Dinh, Husmann & Melloni, 2023). **The EU's Data Governance Act (DGA), on the other hand, aims to incentivise data sharing in order to achieve and promote the single market in data, principally through incentivising 'the voluntary sharing of data between different actors, inter alia through the creation of a legal regime for data intermediaries ('DIs')**, that is to say entities that intermediate between data holders/subjects and potential data users to facilitate sharing of personal and non-personal data' (Carovano & Finck 2023, p.2, emphasis added). The data intermediaries could be private for-profit providers which charge for their services and may be able provide services in addition to mere intermediation, although this remains unclear legally under the DGA (Carovano and Finck 2023). The DGA provides a 'horizontal framework' for data spaces, including the GDDS. As regards the GDDS, Noto La Diega and Derclaye (2023, p.4) consider its development to be 'a step in the right direction' in terms of data being opened up for sustainability purposes, 'but it is questionable whether it provides sufficient incentives for big [technology companies] to share their data' (p.4).

In conclusion, there are various pre-existing examples of voluntary and mandatory B2G data sharing relevant to the GDDS. The new legal frameworks in the Data Act and DGA encompass mandatory data sharing in certain circumstances in the former, and set out mechanisms and frameworks to facilitate voluntary data sharing in the latter. From examples of prior voluntary B2G data sharing, we can discern benefits to the public sector from such access and use of business data, although the incentives and benefits to the private sector of allowing this are less clear. However, this can be overcome for voluntary B2G data sharing by presenting a compelling case for B2G data sharing to address pressing needs. Climate change can be viewed as such, if a clear and possibly specific case is made for the need for business data to address an aspect of climate change. Reciprocity, trust and mutual benefits can assist in leading to B2G data sharing.

However, **it should be noted that not all private sector actors can be persuaded to share data voluntarily, and their willingness to do so may depend on factors such as the size of the recipient government agency**. There are tools in the new Data Act to oblige recalcitrant companies to engage in such data sharing in certain circumstances, and the collaboration of different sized public agencies in data spaces such as the GDDS may also overcome this issue. Further opportunities may arise in the due diligence and compliance required by new legislation such as the Deforestation Regulation, which involve data being gathered

to comply with such obligations[29]. If such companies involved can be persuaded of broader public interest uses of such data for sustainability purposes, then these might form important sources of GDDS data. Yet, if companies cannot be persuaded of the value in sharing this data, the EU may need to consider further mandating B2G data sharing to cover, e.g. the data gathered anyway via these obligations, beyond what is currently the case in the Data Act.

5.2.3 B2B

B2G sharing is not the only form of data flows envisaged by data spaces including the GDDS. Data spaces should include data from a number of companies, and involve data flows among them as well as between them and the public sector. Therefore, business to business (B2B) data sharing is another dynamic within data spaces.

Hitherto in the EU, B2B data transfers have tended to take place on a voluntary basis under the principle of freedom of contract, with non-regulatory measures, e.g. guidance as the main means for facilitating this prior to recent regulatory interventions. There is some compulsory B2B and B2C data sharing where there are or may be market failures which cannot be addressed by competition law (Espinosa Apraez, 2021).

In 2018, the Commission published a Staff Working Document, Guidance on sharing private sector data, which included a set of principles for B2G and B2B data sharing. For B2B contexts in particular, it set out models, suggestions for the content of data sharing agreements, and technical aspects. The 2020 European Strategy for Data (p.7) also mentions B2B data sharing, noting that it had 'not taken off at sufficient scale' and offered the following reasons for this:

> *This is due to a lack of economic incentives (including the fear of losing a competitive edge), lack of trust between economic operators that the data will be used in line with contractual agreements, imbalances in negotiating power, the fear of misappropriation of the data by third parties, and a lack of legal clarity on who can do what with the data (for example for co-created data, in particular IoT data).*

These reasons also hint at economic asymmetries in businesses' knowledge and understanding, which may in turn reflect different sizes and capabilities of different businesses. Just as the public sector is not monolithic and includes different kinds of organisation at different levels and of different sizes, the private sector is also heterogeneous – with small and medium-sized enterprises (SMEs) often in a vastly different position as compared to, e.g. transnational Big Tech companies, also regarding their data literacies and capacities.

Zoboli (2023, pp.13-14) identifies main forms of B2B data sharing:

- **data monetisation**;
- **exchange through trusted intermediaries** providing secure online platforms and receiving payment per transaction;
- **industrial data platforms** whereby 'companies voluntarily decide to establish closed and secure environments to facilitate the development of new products and/or services and/or to improve their internal efficiency';

29. Regulation (EU) 2023/1115 of the European Parliament and of the Council of 31 May 2023 on the making available on the Union market and the export from the Union of certain commodities and products associated with deforestation and forest degradation and repealing Regulation (EU) No 995/2010 PE/82/2022/REV/1 OJ L 150, 9.6.2023, pp.206–247.

- **open data policies** whereby companies share data freely; and
- **technical enablers** – APIs being among the most common – released by specialised companies which obtain 'revenues from the creation, use and/or maintenance of the technical solution, and not from the data exchanged'.

In particular, the role of trusted intermediaries to overcome current reluctance in data sharing might again be crucial in the GDDS context. They are expected to facilitate data exchanges through technical means (from the establishment of specific infrastructure to the offering of tools and services for temporary storage, curation, conversion, anonymisation, etc.) but also legal or other means, to support the formulation and implementation of data sharing agreements, among others.

From the economic and legal perspectives of B2B data sharing, Martens et al (2020) acknowledge the potential gains for welfare and innovation from such sharing and also the potential harms in terms of building and entrenching monopoly positions and weakening incentives for firms to collect data. Studlein (2022) identifies further **economic barriers to B2B data sharing in the forms of unawareness, uncertainty and incapability, alongside legal barriers (liability, ownership of non-personal data, and access and usage rights) and technical barriers (lack of advanced technical infrastructure, technical specifications and standards)**. Zoboli (2020, p.8) supplements this list with further disincentives in the form of needing to safeguard personal data and comply with the GDPR, along with deficiencies in 'ad hoc standards, licensing models and mechanisms for establishing the value of datasets', while considering that legislative measures and policy guidance have contributed to ameliorating the situation for B2B data but was still insufficient to encourage B2B data sharing. She also points to **the use of private initiatives to set standards for data formats and semantics and the issuing of model contracts as helpful in achieving more B2B data sharing**.

Studlein (2022, p.1) considers that **despite the promises of voluntary B2B data sharing, it is 'rarely established' and even when it does happen, it is very selective, limited and mostly relates to customer data**. She mentions as a – rare – example (and one relevant to the GDDS) the mapping service HERE, which was established by competing car manufacturers Daimler, BMW and Audi, and assists with transportation sustainability along with making profit for the companies involved[30].

Martens et al (2020) point to pre-existing forms of mandatory B2B data sharing such as via the essential facilities doctrine in competition law under Art. 102 of the Treaty on the Functioning of the European Union (TFEU), which may give another firm access to a dominant actor's data in certain, limited circumstances (see also Daly, 2016). Furthermore, if access to data is sold by a dominant undertaking, there may be unfair price discrimination at play if downstream operations of the dominant undertaking can access this data for less or no price. However, Zoboli (2023, p.17) considers that the essential facilities doctrine is unlikely to apply as 'data-based markets, however, do not seem to possess the conditions required' for its application, a view also supported by Borgogno and Colangelo (2019). Martens et al (2020) suggest third-party intermediaries, as

30. https://www.here.com

mentioned above, as a way of overcoming the challenges they identify and thereby facilitating B2B data sharing as envisaged in the data spaces. Tombal (2021, p.3), however, points to the limits of voluntary B2B data sharing, noting that 'legal instruments promoting voluntary data sharing, which will tend to focus more on data governance and technical issues (standardisation, interoperability, etc.), in order to create more favourable conditions for the market actors to remedy, or at least reduce, these market failures themselves' (emphasis in original), and advocates for more compulsory B2B data sharing (albeit within limits and balanced against harms).

There are some further forms of B2B data sharing provisions currently in EU law. The Regulation on the free flow of non-personal data[31] addresses 'data sharing practices in the commercial arena (business-to-business)' (Borgogno & Colangelo, 2019, p.3). Among its provisions, Art. 6 creates a B2B data portability right for non-personal data. The Data Act also facilitates some B2B data sharing, tied to B2C data sharing, which is discussed below. The new Digital Markets Act, the ex ante regulation for large online gatekeepers, does contain some compulsory B2B data sharing provisions, 'giving commercial users the possibility to analyze the data directly on the gatekeepers' databases' which could be 'a valuable tool to facilitate B2B data sharing, complementary to data spaces' (Woźniak-Cichuta, 2023, p.7). It is unclear how relevant the DMA data sharing provisions will be to the GDDS specifically. The European Commission has so far designated six gatekeepers – Alphabet, Amazon, Apple, ByteDance, Meta and Microsoft – under the Digital Markets Act. In total, 22 core platform services provided by gatekeepers have been designated. Some among these designated gatekeepers – namely Google and Microsoft – are members of the EU Green Digital Coalition[32], which does suggest they have some interest in green data.

Against this backdrop, the DGA seeks to facilitate more B2B data sharing. On this point, Schneider (2022, p.7) mentions data sharing service providers established under the DGA as being acknowledged by the European Commission as 'a particularly suitable model for business-to-business data sharing, which has been until now left without regulatory coverage'. However, Studlein (2022, p.46) is wary of measures in the DGA to facilitate increased data sharing as:

> *entail[ing] high compliance costs for intermediaries and major uncertainties for companies particularly with regard to a) measures to protect disclosure of non-personal data, b) the GDPR and antitrust law, and c) responsibilities of regulatory authorities.*

Studlein (2022), however, welcomes the data spaces as a way of overcoming some of the technical barriers she identifies, especially as regards data interoperability standards and infrastructure creation.

Overall, there is limited B2B data sharing, most of which is voluntary at the current point in time. As regards mandatory B2B data sharing, there are some specific provisions which oblige B2B data sharing in specific circumstances, leading to a fragmented picture. Problems and obstacles which beset B2B data sharing such as compliance costs, uncertainties, the

31. Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union (Text with EEA relevance.)

PE/53/2018/REV/1 OJ L 303, 28.11.2018, pp.59–68.

32. https://www.greendigitalcoalition.eu/coalition-members/

protection of personal and non-personal data, the application of competition law, and the role of regulatory bodies need clarification for each data space in order that they successfully facilitate B2B data sharing dynamics.

5.2.4 B2C

A further relevant data sharing dynamic is that of business to consumers (B2C). The general public or subsets of publics are envisaged as participating in data spaces; therefore, it is appropriate to consider what current mechanisms exist for transfers to them, usually in their conceptual guise as consumers.

Currently there are several provisions in EU law which facilitate some B2C data sharing, in certain circumstances and/or for certain types of data. There is scepticism concerning how successful these B2C mechanisms have been in practice in terms of consumer take-up and in achieving broader consumer and competition objectives. For instance, regarding personal data, the GDPR contains a limited data portability right for data subjects in Art. 20. However, the data portability right 'so far has not fulfilled the expectations of more competition, more innovation and a solution to lock-in problems through lowering switching costs' (Kerber 2023).

There are also consumer law mechanisms for B2C data transfers in the Digital Content Directive, and some sector-specific tools such as the access to account data in the Second Payment Services Directive (PSD2), the Motor Vehicle Regulation, and the Electricity Directive (Martens et al, 2020). As regards the PSD2's access to account rule, it:

> *aims to address the competitive concerns affecting the retail banking sector and nurture the development of FinTech innovation, stipulates that 'account servicing payment service providers (ASPSPs), such as banks, shall allow third parties to obtain real-time data relating to customers' accounts as well as provide access to such accounts by executing payment orders initiated through digital interfaces, on condition that customers give their explicit consent and that the account is accessible online* (Borgogno and Colangelo, 2019, p.8).

The logic of such pre-existing consumer data mechanisms could facilitate B2C data sharing in the GDDS, although where such sharing is voluntary or only triggered at the request of the consumer, their limited impact needs to be taken into account when designing the GDDS.

Combining B2B and B2C data flows has another precedent in the Data Act's provisions related to the governance of IoT-generated data, which are intended to facilitate increased consumer (B2C) and business (B2B) access to IoT data from device manufacturers and include 'new rights for the users of IoT devices to get access to this data and share it with other firms'. (Kerber 2023, p.120). The rights in Arts. 4 and 5 of the Data Act facilitate users – which is defined broadly in Recital 18 as 'a natural or legal person, such as a business, a consumer or a public sector body' – being able to access generated data (raw data and not derived and inferred data) 'without undue delay, free of charge and, where applicable, continuously and in real-time' (Art. 4.1) – 'and share that data with third parties such as but not limited to other businesses' (and so could potentially include intermediaries and data altruism organisations), which are entitled to receive that data from the data holder 'without undue delay, free of charge to the user, of the same quality as is available to the data holder and, where applicable, continuously and in real-time' (Art. 5.1) . They are not permitted

to share the data with DMA-identified gatekeepers, but could share data for a price to third-party businesses and could do this via providers of digital intermediation services under the DGA (Kerber, 2023).

These new data access rights provided by the Data Act, along with the regulation of data intermediaries in the DGA, set the foundation of the European Strategy for Data (European Commission, 2020) to unlock more availability of data across the economy. Specifically, **data on the environment are often captured by many of these connected devices, enabling companies to enhance their capacity to not only report on their environmental impacts but also develop effective strategies to improve the efficiency of their supply chains and create more accurate sustainability strategies**. However, it is still too early to assess how this new regulatory framework will be implemented, and whether businesses will be able to take advantage of the opportunities it presents.

Recent academic studies provide a critical analysis of the potential limitations that the Data Act may present. Kerber (2023, p.126) identifies various challenges, e.g. 'large obstacles, (transaction) costs (fees, negotiation costs, solving of disputes, technical protection) and delays, which might make this mechanism for third parties potentially very expensive and slow', which cast doubt on how effective Art. 5 of the Data Act will be in facilitating new and better services. Kerber (2024, p.7) also considers that the Data Act does not do enough to redress the imbalance between IoT manufacturers and other businesses on the one hand, and consumers on the other, as the manufacturers can still 'set unilaterally the rules for the IoT ecosystems and for the extraction of the value of IoT data through their technological decisions and contractual arrangements', concluding that 'much more has to be done to give consumers more control over their IoT data (and their IoT devices), and, in particular, to protect and enable more competition and innovation'.

Notwithstanding these criticisms, IoT-generated data may contain 'data related to the environment … that enables indirect conclusions about the environment to be drawn, or to influence the environment', examples of which may be traffic emissions data or farming fertilisation data, which can in turn provide insights on how to address emissions and optimise cultivation (Finck & Mueller, 2023, p.114). However, Finck and Mueller (2023, p.131) also consider that the current frameworks in the Data Act and DGA do not go far enough in facilitating 'access to data for the environment' and instead may require further measures such as:

> *compulsory forms of B2B data sharing, as was discussed by the Impact Assessment for the Data Act but ultimately rejected in the regulatory process. More generally, the availability of environmentally relevant data could also be regulated by a sui generis, sectorial regime, which both the draft Data Act as well as the DGA seem to leave open as an option.*

Further data which may be relevant to the GDDS will be contained in mandatory digital product passports being introduced by the forthcoming Ecodesign for Sustainable Products Regulation[33] which will provide

33. Regulation (EU) 2024/1781 of the European Parliament and of the Council of 13 June 2024 establishing a framework for the setting of ecodesign requirements for sustainable products, amending Directive (EU) 2020/1828 and Regulation (EU) 2023/1542 and repealing Directive 2009/125/ECText with EEA relevance. PE/106/2023/REV/1 OJ L, 2024/1781, 28.6.2024, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1781&qid=1719580391746

sustainability information across a product's entire lifecycle and through its value chains, culminating in information available to consumers. Such compliance data, provided to businesses and consumers along the value chains, could form a source for the GDDS.

5.3 Opportunities and challenges

This section identifies and discusses in more detail recognised opportunities and challenges for B2G, B2B and B2C data sharing within data spaces, with an emphasis on voluntary data sharing by the private sector rather than compulsory data sharing (e.g. via legal obligations). As mentioned earlier, the benefits of such data sharing for the public sector have been recognised for some time, but the benefits for businesses in doing so are less clear or obvious. One of the challenges recognised for the public sector is the very participation of businesses in the first place in such initiatives. Given this is an understudied topic, opportunities from a business perspective on the public good as regards data sharing are considered, drawing on material from corporate social responsibilities, business incentives to access public sector data and environmental reporting. Next, distinct challenges to business data sharing are examined, which mainly emanate from existing legal regimes including intellectual property, confidentiality, and competition law, as well as power imbalances between relevant data space actors and participants.

5.3.1 Opportunities: A business perspective on data and the public good

This section looks at a number of opportunities for business involvement and incentives in participating in data spaces, particularly GDDS. First, profit motives are considered, before proceeding to corporate social responsibility and reputation and trust.

5.3.1.1 Profit incentives

Businesses may share data for profit incentives if they consider that selling or otherwise making data available may directly generate a profit or assist with profit generation. If this is possible within the GDDS, then appealing to this self-interested perspective of businesses may encourage them to participate voluntarily in the data space. If participation in the data space involves accessing certain kinds of data that might not otherwise be accessible to them, from the public sector for instance, which the company considers it can use to generate profit-making outputs, then that would be an incentive to participate in the data space. In some sectors such as health, public-private data sharing partnerships are more common, and can be lucrative for the private sector as participation can involve access to high quality public health datasets from which they can generate private benefit in the form of profits, while producing public benefits as well in the form of new treatments, drugs, products, etc. It is unclear whether there are similar scenarios which arise in other sectors beyond healthcare. The dynamics of healthcare in EU countries exhibits a strong public sector character (in contrast to the US, for example), which means that the public sector has large amounts of high quality, valuable data. The extent to which this can be readily replicated in other sectors is unclear.

5.3.1.2 Corporate social responsibility

Corporate social responsibility (CSR) can be an incentive for businesses to share data for the

public good. As discussed in more detail above, there are various examples of voluntary B2G data sharing in emergency contexts, including those which have an environmental or climate aspect. This could be built upon from a broader CSR perspective. Indeed, Schneider (2022, p.24) considers that 'the sharing of companies' data to the public sector could itself become a CSR-based practice in the very near future and thus be further promoted in consistency with the European Strategy for data's objectives'. Schneider (2022, p.19) recognises the newly emerging notion of 'corporate digital responsibility' as the 'lawful and sustainable management of data and digital technologies', which may entail a much broader public interest role for private sector data and digital technologies. Such a concept could combine CSR on climate and sustainability issues with CSR involving using data and digital technologies ethically to facilitate access to environmentally relevant data held by businesses, and ensure data use is itself carbon neutral, to achieve climate goals in the GDDS.

5.3.1.3 Reputation and trust

Companies can build good reputations and increase their 'trustworthiness' amongst relevant stakeholders by voluntary data sharing. The Science Europe response (2021, p.1) to the European Commission's consultation on the future Data Act mentions 'the soft positive effects that private-to-public data sharing can have for private entities, such as enhancing the entity's public reputation' (quoted in Switzer & Berti Suman, 2023). Schneider (2022, p.22) also sees CSR combined with regulation as being able to 'contribute to building collective trust regarding data collection and use practices vis à vis external stakeholders, with reputational as well as competitive gains for businesses'.

Micheli (2022, p.7) in her study of municipalities found voluntary data sharing was considered part of CSR, but was not perceived by her interviewees as 'philanthropic', instead:

> *it was understood as a marketing strategy private companies used to increase their reputation and to build marketable use cases. Companies share data with cities at no cost – and eventually collaborate with them to develop products or services valuable for the municipality – as a means to develop new data-driven services to be sold to other cities in the future.*

This demonstrates how companies can be incentivised to participate in initial voluntary data sharing due the good reputation and trustworthiness this can generate for them, which in turn can be instrumentalised as making them more profitable and attractive for paid engagements and initiatives with the public sectors such as by winning public tenders to provide digital infrastructure. This can have negative consequences though as recognised by Micheli (2022, p.4): 'through B2G data sharing companies reinforce their position in the market and vis-à-vis public entities'. Such negative consequences can occur when public authorities purchase data from the private sector, thereby generating more income for the companies, which will be problematic if the companies are, e.g. monopolistic or otherwise possess a large degree of market power.

Trust and reputation in the environmental sphere can cut in different ways though, especially given scandals like Dieselgate involving a business falsifying environmental data. Greenwashing must be avoided, which may be aided by the recently adopted EU

Sustainability Reporting Standards. Schneider (2022, p.15) recognises specific accountability issues in the environmental sphere:

> *Although it is true that large companies and groups bear specific disclosure duties regarding non-financial information as environmental information under Dir. 2014/95/EU100, dramatic cases as the Dieselgate have well highlighted the problems regarding the implementation of adequate accountability safeguards regarding companies' generation and management of non-personal information as environmental information.*

Both CSR and reputation and trust offer incentives for businesses to provide data which may be 'legalised' in the form of the new sustainability and circular economy measures that the EU is implementing. These include the deforestation-free products and digital product passports, which will require the provision of data for compliance and due diligence and but not only be subject to businesses' own voluntary ethics.

5.3.2 Challenges

There are various challenges in place inhibiting business sharing of data, many of which are referred to above. Klein and Verhurst (2017) identify various risks and potential harms of B2G data sharing:

> *Risks are often the result of technological weaknesses (e.g. security flaws); individual and institutional norms and standards of quality (e.g. weak scientific rigor in analysis); legal confusion or gaps (e.g. weak or no privacy provisions); or misaligned business and other incentives (e.g. companies seeking to push the boundaries of what is socially appropriate)* (Klein & Verhurst, 2017, p.15).

Klein and Verhurst (2017, p.17) also point to the risks for public trust overall, concerning both private companies providing data and in public organisations using the data, if erroneous or faulty data are involved. Mitigation methods, they suggest, include 'limiting data access to specific, pre-approved uses' and 'bring the algorithm to the data', an arrangement in which private-sector data sets never leave corporate databases but are processed and analysed using external algorithms that may, for instance, be hosted in the cloud' (Klein and Verhurst, 2017, p.15).

Again, the Data Portal Analytical Report (2020) echoes the aforementioned challenges and identifies others, such as the cost and benefits of a B2G initiative including the need for management and curation by a company for data to be ready for onward sharing, and the need for human capabilities and organisational cultures to facilitate data sharing. Data intermediaries and well-designed data spaces may be able to provide such curation services and facilitate sharing among a wide range of players if such risks can be addressed.

A few of these challenges are examined in the following, in particular intellectual property and competition law. Legal issues relate to the broader theme of liability, which is identified in the Data Portal Report (2020, p.14) as contributing to uncertainty risks as to responsibility 'when bad quality data is shared and contributes to a wrong decision by a public body, possibly damaging citizens or causing financial loss'. Mechanisms which clearly address these in data spaces are required from the outset. These mechanisms, and the ongoing operation of the data space, also need to acknowledge the power differentials at play among the different players involved in the GDDS, especially the power held by at

least some private sector actors, to ensure the GDDS achieves its objectives of combatting climate change in the public interest.

5.3.2.1 Intellectual property

Intellectual property (IP) is identified by various authors as a barrier to data sharing. The Data Portal reports considers:

> *Companies may not have the necessary certainty of having the right to share the data they process every day. Company data is often the combination of very heterogeneous sources, from historical archives whose origin is no more known, to datasets that are the results of working with suppliers and contractors, whose relationships may not have been regulated by contracts that were explicit as of the intellectual property of the data produced* (Data Portal Analytical Report, 2020 p.14).

Schneider (2022, p.10) identifies how businesses' IP protects 'both data and the criteria governing their processing', which may lead to restrictions imposed vis-à-vis onward uses of these data by governments in contracts between the businesses and governments, giving rise to public accountability and transparency concerns if that data is used in government decision-making. Noto La Diega and Derclaye (2024, p.15) point specifically to the database right as 'pivotal to opening up data', but identify various ambiguities in the right's application to the kinds of data generated by new digital technologies and applications and in the scope of its exceptions, which are not fully remedied and clarified by the Data Act.

5.3.2.2 Competition law

Compliance with competition law, especially the prohibition on cartels and anti-competitive agreements in Art. 101 TFEU, is also mentioned by many authors as a barrier to more data sharing, especially in B2B contexts – but the existence of monopolies too (another target of competition law enforcement) is mentioned as a barrier to B2G data sharing as well (see e.g. Monahan, 2020).

Woźniak-Cichuta (2023, p.7) considers that data spaces 'should be compatible i.a. with the Horizontal Block Exemption Regulations on Research & Development and Specialization agreements, as well as the Guidelines on the applicability of Art. 101 of the TFEU to horizontal co-operation agreements, which provide rules for the sharing of know-how and information exchange between businesses'. Nevertheless, Woźniak-Cichuta (2023, p.7) points to some of the opportunities of data spaces for competition: they may 'reduce high concentration on digital markets stemming from emerging power of digital conglomerates' and reduce data as an entry barrier.

Competition issues – both in terms of anti-competitive agreements and the role of highly powerful companies – need to be clarified from the outset in data spaces including the GDDS to give clarity and certainty to potential business participants, and to ensure competition is preserved and promoted.

5.3.2.3 Power differentials

A combination of regulation and corporate social responsibility can incentivise businesses to share data with the public sector, and regulation and profit-incentives can incentivise businesses to share data with other businesses. Companies tend to share data with consumers when prompted by regulation. However, these dynamics are imbued with power differentials. Companies with a large amount of data are generally more powerful than other companies and consumers, and the

public-private power dynamic may depend on the specific actors, context, or scenario (see e.g. Chignard and Glatron, 2023).

Some existing laws such as competition law may be considered by some to address, partially, some of these power differentials. However, other laws such as IP may further reinforce them, as Noto La Diega and Derclaye (2023) identify in the accumulations of power that big tech companies possess over data and algorithms, which are created and reinforced by IP, especially database rights and trade secrets. To address this, they also urge a 'holistic' appreciation of data laws and IP in the EU. Even the new Data Act has come under criticism by Kerber (2024, p.7) for being 'unclear' about whether it 'contributes to the containment of the data power of gatekeeper firms', as third-party businesses are able to share users' IoT data with gatekeepers (although users themselves are not). He has previously criticised the Data Act for having the potential to 'lead to the emergence of specialized large data companies who build up entire portfolios of data streams from different IoT devices, combine them (also with other data) and extract value from these huge sets of data' which in turn can 'lead to entirely new forms of data concentration and data power in the digital economy' (Kerber 2023, p.130).

Even if government organisations may be larger and, due to state power, more balanced in terms of their position with some large companies, there are still shortcomings. Schneider (2022, p.11) points to the lack of technical expertise in governments – and also among citizens – as compared to many businesses vis-à-vis the datasets and processing history, which may in turn 'impair the generation of adequate explanatory justifications regarding the process and the reasons that have led the public authority to take a certain decision' if based on the use of private data. This in turn can have further negative effects: 'Governments' loss of accountability in B2G data sharing thus ends up exacerbating public disempowerment'. Not only might there be a loss of government accountability vis-à-vis the public, but Micheli (2022, p.4) recognises:

> *Even when data is offered for 'free' and in an altruistic manner, data sharing can generate power imbalances and dependencies. For instance, national governments in lower-income countries are increasingly relying on private-sector data, often offered at no cost by Telcos and big tech corporations, to gain information about social and economic trends. This has implications for sovereignty since it generates new visibilities and 'states' are not – or are only partially – in charge of collecting and disseminating (...) representations of themselves.*

In Micheli's study of municipalities (2022, p.14):

> *power relations embedded in B2G data sharing bringing to the fore three main themes: the contextual factors that shape access to private sector data and might generate inequalities between municipalities; the willingness of local administrations to actively engage and control how data is shared with them; and the potential of collective efforts to increase local administrations' negotiating-power and address some of the asymmetries of current data ecosystems.*

Again, some of these issues may be addressed by well-designed data spaces which involve

public sector actors joining together to negotiate pools of data with the private sector.

The role and empowerment of the public is also key, especially as the public is not synonymous with state power. In fact, the public and private sectors represent 'elite' positions which may be detached from the public at large, and this can be reinforced by data sharing between them:

> *although there are obvious potential public benefits to B2G data sharing, it should be noted that some data sharing between large companies and governments serves to reinforce corporate and state power over ordinary people. In other words, B2G data sharing takes place between entities (business and governments) whose interests are not necessarily aligned with those of citizens, especially in neoliberal times; rather, they follow elite interests. Therefore, B2G is no panacea and must follow democratic principles in order to achieve its declared inclusive objectives* (Mercille, 2021, p.2).

Accordingly, in designing the GDDS, these power dynamics and differentials must be understood and addressed. Weakest among the actors is the general public, and for the GDDS to work for both people and the planet, genuine public interests and benefits must be produced by it. There is the possibility for well-designed and well-supported data spaces, altruism organisations, and personal intermediaries to address some of these imbalances and work in the public interest.

5.4 Lessons learned and recommendations

From the preceding discussion, there are a number of lessons and recommendations.

Opening up more data from the private sector can be helpful for achieving GDDS goals, and there are some mechanisms and incentives to do this which are elaborated above. Beyond these, the following recommendations are made:

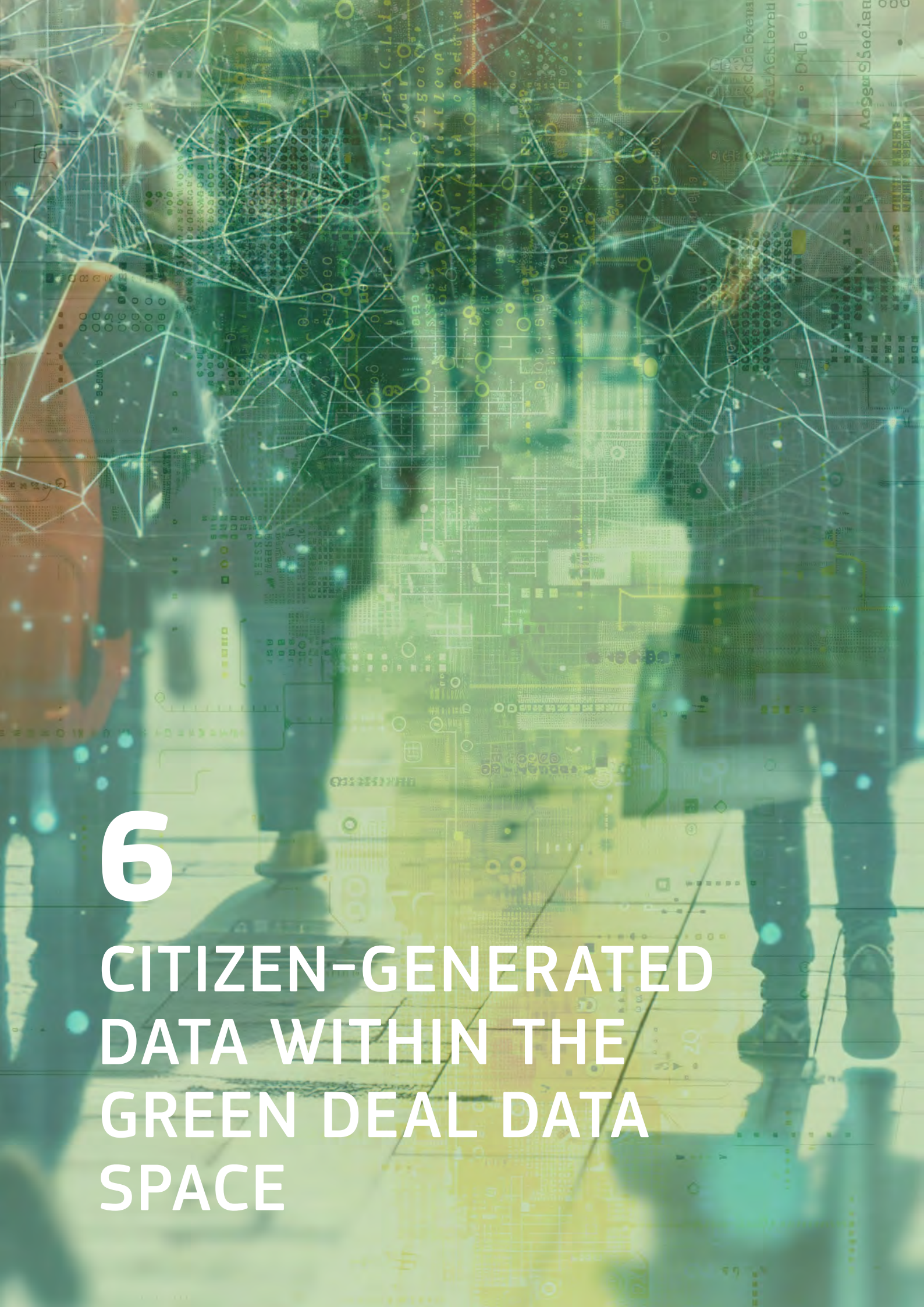
- **Legal certainty would be needed to facilitate business involvement in the GDDS.** This requires clear rules to manage the GDDS from the outset, which should address issues involving liability, the relationship between the GDDS and other areas of law such as competition law, data protection, and IP, and what can and cannot be done with the GDDS data.
- **Accountability as to how data are shared in the GDDS** is also important, along with what happens downstream, especially if data are used in public decision-making, as high-lighted by Schneider (2022), in particular data originating in the private sector.
- **Corporate social responsibility, possibly leading to corporate digital responsibility**, can be leveraged to encourage more business sharing of data, especially in situations of urgency in the public interest. Climate change is increasingly recognised as a societal crisis requiring all measures possible to address it. Communicating this urgency to the private sector and demonstrating how access to their data for the GDDS can help fight climate change can align with their CSR objectives on climate and data/digital technologies. If businesses can be shown that, e.g. the data they collect for sustainability reporting can have important secondary uses in the GDDS, this might form an important source of data sharing, done on the basis

of CSR, or may also give rise to new business models.

- **Recognising and addressing power differentials** is also key to creating a genuinely successful and equitable data space. This must take account of the self-interest of businesses even when they are sharing data voluntarily and the imbalances of skills between business and others. Furthermore, it must understand the genuine, non-elitist public interest and ensuring that is what guides the GDDS.
- **Considering a combination of both voluntary and mandatory sharing of environmentally relevant data, which could also be regulated by a sectorial regime.** Currently the GDDS is mainly reliant on businesses sharing data voluntarily. A better understanding is needed of how this can be done through new and existing business models for the GDDS. In addition, the notion of high-value datasets could be expanded to privately held data to determine which could be made available for broader/compulsory reuse and/or reuse across sectors via a data space such as the GDDS (Espinosa Apraez, 2021). This may be something to consider if, despite appeals to the urgency of climate change and business CSR, there is insufficient access to privately held data once the GDDS has been formed.
- **More business data are expected to be made available through the new access rights provided by the Data Act.** However, further understanding of the capacities of (small and medium-sized) businesses to exercise these rights is needed, as well as the potential role of data intermediaries to overcome data access and sharing barriers.
- **Recognising too that more access, sharing and use data does not necessarily equal better data and better outcomes** is also crucial. A data-determinist approach is risked by making this assumption, which is all the more problematic given the environmental costs of data – perhaps a notion of 'data sobriety' can assist with ensuring only useful and necessary data is shared (see Shulz et al, 2024).

5.5 Conclusions

This chapter has addressed business data sharing as relevant to the GDDS, by identifying different directions of such sharing – B2G, B2B and B2C – relevant to how businesses might share data in data spaces and with whom. In doing so, existing EU laws and policies in this area have been identified and discussed, along with voluntary data sharing beyond legal obligations and its dynamics with respect to opportunities and challenges. Various recommendations have been offered based on lessons learned from the preceding material for the GDDS. In short, appealing to businesses' CSR and the urgency of climate change may facilitate more voluntary data sharing by businesses in the GDDS. Clear rules governing the data space are also required from the outset to provide certainty to businesses which may wish to participate, on issues such as competition, IP and confidentiality. If these measures are insufficient to facilitate the optimal levels of business data sharing, the EU may consider further obligations on the private sector to share, on a mandatory basis, data needed for the successful achievement of the GDDS's goals.



6

CITIZEN-GENERATED DATA WITHIN THE GREEN DEAL DATA SPACE

We are at a turning point in history. On the one hand, disruptive digital innovation and the rapid growth in computing power have resulted in unprecedented data capabilities. On the other hand, environmental depletion is at its highest, and the consequences – from the climate crisis to land degradation – touch all dimensions of people's lives. This chapter explores how the positive consequences of the 'data revolution' can be harnessed for the ecological transition. It examines how individuals and communities, including laypersons, can generate and (re)use data on environmental issues of concern to drive changes in social norms, laws, and regulations and inform policymaking.

In today's increasingly data-driven society, a vast array of tracking and sensing technologies is transforming human and social activities into data points. Never in history has humanity possessed such extensive knowledge about itself and its environment (Cukier and Mayer-Schoenberger, 2013). Data production and analysis are enabled by software and computing infrastructure, including cloud capabilities and data centres. Well-resourced industry players take the lion's share in the data race, operating at a large scale for profit-driven purposes such as the monetisation of personal data for marketing and individual profiling (Couldry and Mejias, 2019). Their activities have a staggering environmental footprint (Monserrate, 2022).

Countering these trends, a distinctive type of data has emerged – citizen-generated data (CGD). Thriving on the fringes of the mainstream data ecosystem, CGD benefit from the recent availability of accessible, low-cost tools, and is often the result of participatory collaborative initiatives. Rather than 'big data', CGD are frequently 'just good enough data' (Gabrys, Pritchard, and Barratt, 2016, p.2), and challenge 'standard practices used by regulators' (Ottinger, 2010, p.245). They provide valuable opportunities to enhance environmental governance and foster citizen participation and empowerment in the pursuit of sustainability. Consequently, CGD can play a significant role within the framework of the EU Green Deal data space (GDDS), in connection with the opportunities and tools provided by the new legal instruments underpinning the GDDS. For example, the definition in the Data Governance Act (DGA) of trusted data intermediaries (European Union, 2022) paves the way to empower citizens' use and management of their own data, and the creation of personal data spaces (Lähteenoja, 2023). Furthermore, the Data Act (European Union, 2023) provides additional data access rights to users of connected devices and services, such as smart watches, home supplies, etc. Lastly, the Open Data Directive (European Union, 2019) makes public sector data freely available for reuse, opening the opportunity for citizens to access and use these data.

This chapter delves into the nature, challenges and opportunities of CGD in the context of GDDS. It has five objectives:

1. Providing an actionable definition of CGD within GDDS.
2. Positioning citizens as knowledge makers, and knowledge held, generated, or supplied by citizens as complementary to academic and professional expertise, and on equal footing with it.
3. Identifying the main challenges encountered by individuals and groups engaging in data generation and data use in the context of CGD in environmental governance.

4. Reviewing policy measures to enable, foster, and support the generation and use of CGD.
5. Providing a set of policy recommendations ('necessary measures') to create an enabling environment for CGD in the context of GDDS.

The chapter is organised as follows. Firstly, it presents the methodology underpinning the research. Secondly, it reviews the conditions that enabled the emergence and growing popularity of CGD. Thirdly, the Findings section opens with an actionable definition of CGD and recognises citizens as legitimate knowers, essential for maximising the potential of CGD in environmental governance. It then offers an action-based concept map to enable policymakers and CGD actors to classify the core activities and phases of the CGD lifecycle. The chapter reflects on challenges and open questions and explores the features of an enabling ecosystem for CGD. Finally, it proposes a set of necessary policy measures to support CGD supply and utilisation, leveraging its potential for the benefit of society as a whole.

6.1 Methodology and data

The chapter answers three main questions:

- What are CGD comprised of and how can such data contribute to promoting environmental awareness and a just, sustainable transition?
- What challenges are encountered by CGD?
- What does an enabling environment for CGD look like? What measures can be kick-started or scaled up, and how can they be encouraged?

Data for this chapter are derived from a qualitative analysis, capable of generating 'thick' evidence (Geertz, 1973) on the problem at hand:

- **Desk research:** This includes academic literature across disciplines, as well as policy papers and grey literature by stakeholders like industry and civil society organisations.
- **Interview and ethnographic data:** Collected within the framework of the DATACTIVE project (2015-2021, PI S. Milan, European Research Council StG no. 639379, data-activism.net).

The term 'citizens' in this context refers to individuals and communities who are not professional data producers or analysts (e.g. laypersons) and/or do not have a codified role in data handling (e.g. those outside the realm of the State and industry). In the context of CGD, the reference to citizenship is detached from its legal definition as the bond between an individual and their State. Here, 'citizens' is a synonymous for 'people', also including those who do not hold citizenship rights.

Finally, this chapter adopts a global perspective to identify commonalities in CGD across diverse contexts. Despite minor socio-cultural differences, the nature and dynamics of CGD are primarily determined by the combined effects of data and data infrastructure on the one hand, and an enabling environment, on the other. These are the primary concerns of this report.

6.2 Background

As society increasingly turns into data ('datafies') various aspects of human activity – from personal health and interpersonal connections to public safety and social security – citizens and institutions alike are becoming more aware of the crucial role of information in democratic systems (Hintz, Dencik, and Wahl-Jorgensen, 2018). This 'datafied society' represents a fundamental paradigm shift in

how we understand and manage the policy and the public good (Kitchin, 2014). However, it is characterised by an unparalleled power asymmetry between the State and its citizens (Brunton and Nissenbaum, 2015). The advance of generative artificial intelligence (AI) augments this power asymmetry.

While today's data-driven environment is often linked to disempowerment rather than democratic agency (cf., Couldry and Powell, 2014), transformative experiments have emerged where citizens repurpose data in innovative ways. *InfoAmazonia* (infoamazonia.org) is one of many examples showcasing the unprecedented potential for laypersons to drive social change through active engagement with data generation and analysis. Since 2012, this network of citizens and journalists from the eight countries of the endangered Amazon region has been leveraging crowdsourced, remote sensing and satellite data to produce investigative reports, maps, and advocacy material. Promoting data transparency in the Amazon through CGD helps to prevent deforestation and wildfires (Milan and Gutiérrez, 2017).

Data from such initiatives are referred to as citizen-generated data. CGD can enable individuals to exercise political agency in a data-driven society (see Milan, 2018). It includes both data that are passively contributed (e.g. cellular phone tracking) and data that are actively generated (Haklay, 2013). When actively generated, CGD often result from a form of political engagement that uses data and data generation techniques as opportunities to drive policy change or societal transformation, known as 'proactive data activism' (Milan and van der Velden, 2016).

The origins of CGD trace back to citizen science initiatives and align with the role of technological innovation in democratising data collection processes. Early manifestations of CGD were often manual and localised, but the penetration of user-friendly digital tools in society, such as smartphones and low-cost sensors, has expanded the scope and scale of data collection efforts. Accessible data analysis and data visualisation software have further facilitated this expansion (Gutierrez, 2018). Interestingly, the affordances of a given technology – the possibilities and limitations it offers to its users (Davis and Chouinard, 2016) – shape how individuals can interact with such technology and what actions are possible, also in relation to the users' abilities and goals, and mould the societal dynamics it triggers (cf., Baack, 2018).

The growing interest in CGD goes hand in hand with the evolving societal perception of what constitutes legitimate knowledge at a given point in time and how it is achieved. Gray, quoted in Kitchin (2014), argues that science has entered a fourth paradigm, predicated on the increased availability of new data and new analytics. This fourth paradigm is a type of 'exploratory science' that is data-intensive and rooted in statistical exploration and data mining. Haklay et al. present citizen science, which involves non-experts in data generation, as a valid tool in 'post-normal science' – a framework 'concerned with the social robustness of applied science, science-based professional consultancy, and scientific advice for policy in situations of high stakes, high uncertainties, and contested values' (Haklay et al., 2023, p.1). Amidst these developments, **CGD can be seen the manifestation of a new way of engaging in knowledge generation**. Its disruptive potential has reconfigured how evidence is produced within the realm of participatory governance.

In the context of environmental and climate governance, CGD are frequently mobilised

as part of environmental justice initiatives, especially when they include contestation over knowledge claims from the ground-up (Ottinger, 2024). CGD can contribute to 'environmental democracy, a set of principles and practices that allow people to participate in environmental decision-making processes that affect their lives' (Berti Suman et al., 2023). Environmental democracy was first enshrined in the Convention on *Access to Information, Public Participation in Decision-Making, and Access to Justice in Environmental Matters* (also known as the Aarhus Convention, 1998) adopted by the United Nations Economic Commission for Europe (UNECE). It is grounded on three key pillars deemed essential for empowering people and communities to protect their environment and to hold governments and other entities accountable for their actions affecting the environment:

- Openness and transparency in environmental governance, including access to information.
- Participation by the public in environmental matters of concern.
- Access to justice in environmental matters and decision-making.

The Commission has also addressed the importance of CGD for the environment, as highlighted in the 2020 Staff Working Document on 'Best Practices in Citizen Science for Environmental Monitoring', which provides key insights on the opportunities for and benefits of using citizen science for environmental monitoring (European Commission, 2020b). Examples of CGD deployed within environmental justice initiatives include a bottom-up initiative monitoring the noise levels of Amsterdam Schiphol airport (Netherlands), where citizens autonomously measured noise levels of manoeuvring airplanes with microphones to challenge the 'informational monopoly' of the State and reclaim their right to live in a healthy environment (Berti Suman, 2018). Similarly, in the US, in oil fields subjected to 'fracking' (an invasive drilling technique used for extracting oil or natural gas from deep underground), professional researchers have developed monitoring kits for residents to track air pollution, one of the byproducts of fracking, thus evidencing harm (Gabrys, 2017). In the riverbed of the Rio Magdalena, one of Colombia's largest watercourses, a group of fishermen concerned about the impact of large-scale gold mining created a time-series of riverbed transformation, through drawing on paper and large canvases. The effort was aimed at advocacy in the absence of updated satellite imagery (author's fieldwork, 2017). But people may also mobilise to defend government-generated data at risk of deletion. Following the 2016 election of climate-denier Donald Trump to the US presidency, citizens organised 'DataRescues', harvesting, scraping and archiving federal websites and datasets related to climate change in fear that they would be removed (Currie and Paris, 2018). It was a way to engage in 'data resistance' (Vera et al., 2018).

These cases demonstrate the growing public interest in CGD to support environmental knowledge and policymaking, acknowledging its potential to 'contribute to the framing of or offer a different perspective on an environmental issue they care about' (Berti Suman et al., 2023, p.2). Despite the democratic potential of these citizen-led responses to the environmental crisis, CGD faces obstacles when trying to give teeth to the knowledge it generates. One such obstacle concerns the recognition of the legitimacy of citizen-generated knowledge. This is because citizen science approaches 'often do not fit within current institutional practices' prevalent in policy cycles (Schade et al., 2021, p.361). Scientists'

preferences might also hinder the adoption and limit the applications of citizen science data, due to the scarce awareness of citizen science projects and dynamics, a persisting distrust in citizen science data quality, and a bias favouring data collected by fellow scientists (Burgess et al., 2017). Moreover, people from marginalised communities or identities typically suffer from systematic 'exclusion from judgment, inadequate epistemic resources, and denial of status as knowers' (Ottinger, 2024, p.199).

Effective promotion and deployment of CGD depends on society's ability to recognise it as a valid form of knowledge generation. A renewed lexicon is needed, which is presented next.

6.3 Findings

6.3.1 Key terms and definitions in citizen-generated data

This section proposes a renewed lexicon for CGD, assuming that the way we imagine and discuss the nature and role of CGD influences their likelihood to impact societal transformation and policymaking. First, it provides an actionable notion of CGD. Second, it posits individuals and communities as knowledge makers capable to accelerating the ecological transition.

BOX 3.

Actionable definition of CGD.

CGD indicate the diverse range of information that:

- is created or collected, analysed, reused, and/or disseminated by non-specialists, including both individuals and community groups or thematic collectives (i.e. mobilising on a specific matter of concern);
- independently or in collaboration with specialists, including scientists or subject-matter experts as well as governmental and non-governmental organisations;
- often but not necessarily leveraging digital technologies;
- either not-for-profit, or in combination with for-profit activities; and
- adheres to ethical guidelines and privacy protections to ensure responsible and respectful data use.

CGD are generally ascribed to the realm of 'non-traditional data sources' (Fritz et al., 2019). Two sets of descriptors provide a structured way to define and understand the various aspects and applications of CGD. These descriptors are not mutually exclusive.

In relation to the nature of CGD, we can identify five sets of descriptors, whose combination allows for labeling information generated by citizens in the context of GDDS:

- **Data types:** Quantitative data (numerical data such as counts, measurements, statistics), qualitative data (descriptive information in text, photos, videos, audio recordings), spatial data (georeferenced data, localised data), temporal data (capturing changes over time).
- **Data collection methods:** Sensor-based data (generated using sensors to provide measurements), crowdsourced data (information collected by many

contributors via platforms or apps), remote sensing data (generated using drones or other remote sensing technology), manual observations (through human scrutiny such as reporting, journaling, etc.).

- **Tempo:** Real-time data, historical data, time-series, etc.
- **Engagement level:** Individual, community or collective, collaborative (leveraging multiple sources of observation and expertise); passively generated data.
- **Data quality:** Raw data (unprocessed information), processed data (cleaned, validated, and analysed), or verified data (cross-checked for reliability).

Citizen-generated data can contribute to the understanding, monitoring, and management of environmental issues. Simultaneously, it typically aims to promote transparency, inform policy, and/or foster community engagement. In relation to its purpose and use, we can identify four main orientations in CGD:

- **Policy-oriented:** When data availability and policy formulation are enhanced by CGD. For instance, CGD can supplement official data and support policy development and implementation.
- **Process-oriented:** When transparency is enhanced by CGD. For example, CGD can promote transparency and accountability in environmental decision-making.
- **Outcome-oriented:** When adaptive management is enhanced by CGD. For example, CGD can improve the responsiveness and adaptability of institutions and foster community resilience.
- **Public-oriented:** When public engagement is enhanced by CGD, encouraging awareness raising, promoting literacy, and boosting public participation, thus empowering citizens to actively engage in environmental governance.

6.3.1.1 Citizens as knowledge makers

Individuals and groups are positioned by CGD as knowledge makers, capable of highlighting on-the-ground realities that might otherwise be overlooked by institutional efforts. They can identify and expose the community's experience of a given problem, including quantified measurements as well as actionable evidence on perceptions, fears, and needs. In doing so, they leverage and become a proxy for a type of expertise derived from lived experiences rather than academic or professional qualifications: **experiential knowledge**. Experiential knowledge can complement academic and professional expertise, offering a holistic view of complex issues and contributing to more inclusive and effective policymaking.

The recognition of citizens as knowledge makers is supported by a growing body of literature, particularly within science and technology studies (STS). Concepts such as 'boundary work' (Gieryn, 1983), 'co-production' (Jasanoff, 2004), 'street science' (Corburn, 2005), and 'civic technoscience' (Wylie et al., 2014) have emerged to problematise the sources of scientific expertise and underscore the importance of experiential knowledge as a legitimate and powerful form of expertise. However, experiential knowledge often lacks recognition in policy cycles or society (Ottinger, 2010). The problem originates from how the category of 'expert' is defined and the public trust attributed to it. As a result, 'alternative, more culturally rooted and legitimate forms of collective, public knowledge – and of

corresponding public order – which could arise from the informal non-expert public domain are inadvertently but still systematically suppressed' (Wynne, 1998, p.46). This may be ascribed to the tension between expertise and experience, which can be addressed by distinguishing between distinct types of expertise, including 'interactional expertise', whereby laypersons gain sufficient understanding of an issue to meaningfully engage with scientific discourse (Collins and Evans, 2002). Importantly, knowledge produced by citizens is 'not only local, situated and experiential in origin but also *collectively generated and held*' (Scott, 2016, p.261 emphasis added).

The sociology of social movements has further emphasised this collective dimension of knowledge-making. Grassroots movement actors can create 'collective spaces of knowledge production' that 'foster the coordination of disconnected, local, and highly personal experiences and rationalities within a shared cognitive system', supporting collective claim-making (Della Porta and Pavan, 2017, p.297). Movements are the harbingers of subaltern knowledge, opposed to official knowledge, where the standpoint from which knowledge is elaborated matters (Cox and Flesher Fominaya, 2009). They focus on power inequalities, the realities of vulnerable communities, and the role of reflexivity in the making of knowledge (Ryan et al., 2010).

A data justice perspective, explicitly integrating social justice concerns into the analysis of data-driven societies (Dencik and Sanchez-Monedero, 2022), stresses the meaningful participation in knowledge generation of all stakeholders, especially those impacted by data practices. Taylor (2017) identifies three pillars of data justice: visibility, engagement with technology, and non-discrimination. These pillars highlight crucial elements such as access to representation, autonomy in technology choices, and the capacity to challenge data biases, which speak to the role of citizens as knowledge producers. In environmental governance, this approach helps counter the 'extractive logic' of the State when it separates data from its provenance (Vera et al., 2019).

The question that remains largely unaddressed in these accounts is that of legitimacy, which can be defined as the condition whereby an actor is regarded by others as exercising its knowledge-making power in an authoritative and appropriate manner (Berti Suman et al., 2020). Legitimacy is not just contingent backing, where citizens as knowers are granted limited, ephemeral support in certain policymaking processes. At a more fundamental level, it involves trust (in the quality of the knowledge produced), and confidence (in the dynamics of knowledge production) (adapted from Scholte, 2019). Acknowledging the role of individuals and communities, including laypersons, as knowers means reclaiming and reinstating the legitimacy derived from experiential rather than merely credentialed knowledge. The next section explores the lifecycle of CGD, which contributes to elucidate why citizens are legitimate knowers.

6.3.2 The lifecycle of CGD: An action-based concept map

Analysing the ideal-type lifecycle of CGD is crucial to understanding the dynamics of CGD generation and use. The lifecycle of CGD includes six stages, illustrated in Figure 1, which presents an 'action-based concept map' representing the relationships between the stage of CGD, with each stage being described by an action verb. This is in line with the notion of **data as enabler** and CGD as enabling a set of activities (namely, activism, advocacy,

solution development), but also transformations (citizen empowerment, behavioural change, and citizen agency). The map is illustrated with examples from the realm of environmental and climate governance.

There are two modes of **making data**, depicted on the left-hand side of the map. Citizens can

- **Collect existing data**. Generated by utility meters, smartphones, or meteorological stations in an automated fashion, data are assembled, analysed and deployed by citizen knowers for awareness raising, evidence generation, or advocacy. For example, the EU-funded DAIAD project (2014-2017) empowered urban dwellers to monitor their data consumption by collecting readily available records to be compared with consumption figures of reference groups, with the goal of lowering water use (Sartorius, 2017). The DesPat app turns Android smartphones into a privacy-respecting camera-based pedestrian tracking tool to analyse pedestrian traffic patterns in an automated fashion (Getschmann and Echtler, 2021). But citizens can also
- **Produce new data** about matters of concern for which evidence is not yet available. The EU-funded D-NOSES project (2028-2021, dnoses.eu) is a case in point: to tackle the problem of odours, it empowered citizens to map and report odour incidents in real-time, notifying type and intensity in the *OdourCollect* app. It gathered over 10,000 odour observations from 1,600 registered users. *FreshWaterWatch* (freshwaterwatch.org) involves individuals and communities in the collection of quality data. 'WaterBlitz' events, organised at regular intervals, enable swaths of citizens to simultaneously test their local waterbody with freely distributed water testing kits.

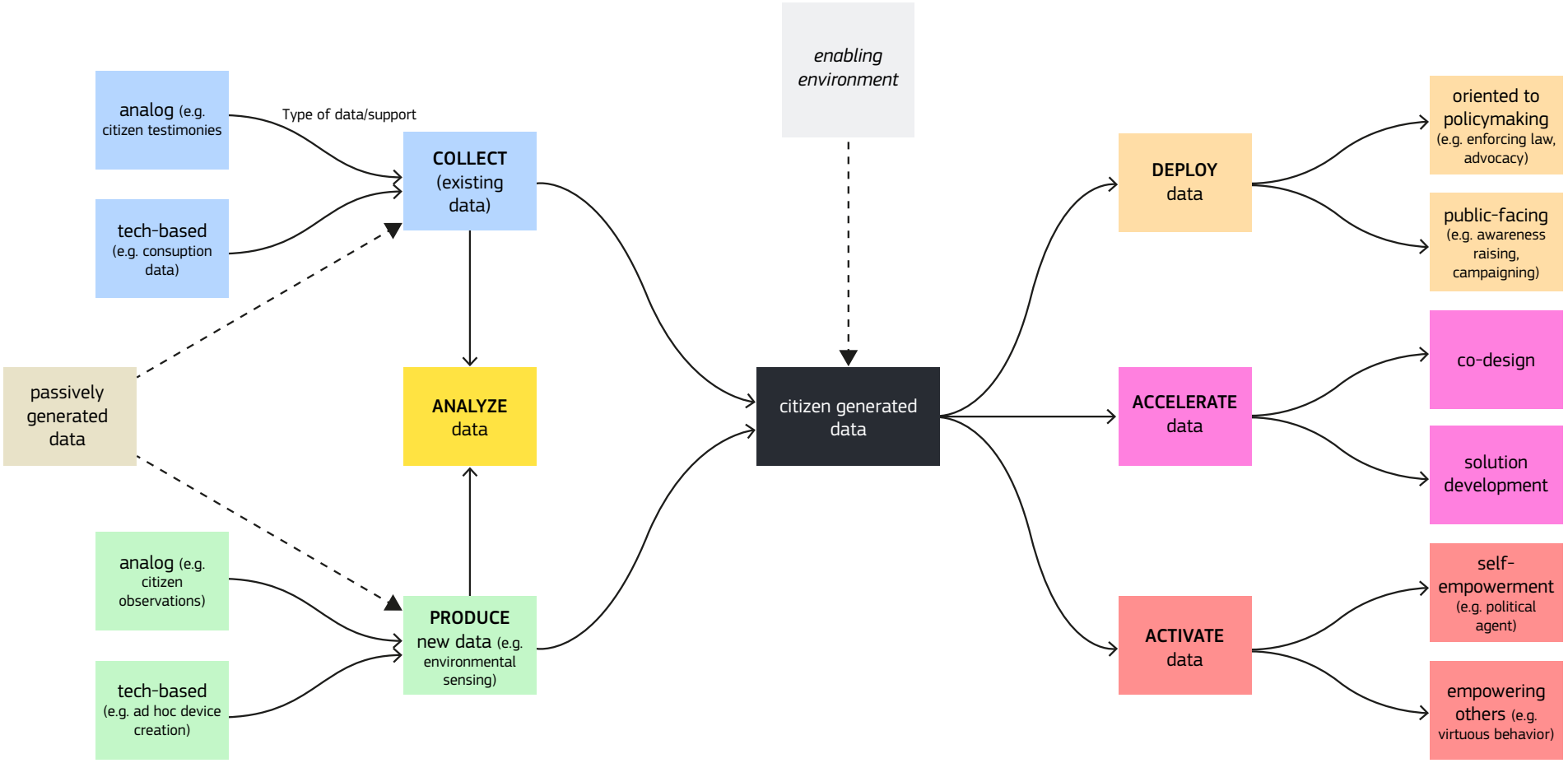
In both instances, we can differentiate between analogue data (such as citizen testimonies or logs and diaries commonly used in conservation projects) and tech-based data, including various digital and non-digital modalities. Examples of tech-based data include ad-hoc devices for environmental sensing or curated data from buckets used to monitor air toxics (Ottinger, 2010). Both analogue and tech-based data can also incorporate passively generated data.

On the right-hand side of Figure 1, we find three main modes of **exploiting CGD**. Citizens can

- **Deploy data:** CGD are curated and arranged for consumption by third parties (e.g. State, industry, public), either mobilised for informing policymaking, enforcing laws or institutional change, or for awareness raising, campaigning, and public education. This is the most popular way of leveraging CGD. For example, people monitored air pollution near fracking sites to make visible the 'experience of citizens living on the gas fields' and seek redress (Gabrys, 2017, p.177).
- **Accelerate data:** CGD are used for co-design (e.g. when citizens model policies jointly with public officials) or solution development (including product development). In Spain, the D-NOSES project accelerated citizen data to develop a national odour pollution standard (2017).
- **Activate data:** CGD are used to enhance individual practices or empower others by promoting virtuous behaviour. This approach is particularly significant in achieving the Green Deal objectives, as it encourages citizens to improve their daily actions. The EU-funded Waste4Think project (2016-2020) use data to illustrate campaigns aimed at reducing landfill waste and increasing recycling in various European towns (Waste4Think, 2020).

FIGURE 4.

Citizen-generated data: An action map.



Source: own elaboration.

6.3.3 CGD: Challenges and open questions

Six pressing issues represent the main weak spots in the CGD lifecycle, and might hamper the validity, legitimacy, and utilisation of CGD in environmental and climate governance. They combine public-facing and internal aspects of the CGD lifecycle.

- **Perceived data quality**, as defined by scientific norms, is often seen as the main obstacle to the legitimacy and validity of CGD (Fritz et al., 2019). The term holds different meanings for various stakeholders (Balázs et al., 2021), while policymakers require 'regulatory-quality data' (Ottinger, 2024, p.207). Moreover, data quality requirements often clash with the open, collective, and anonymous nature of CGD (Lukyanenko, Wiggins, and Rosser, 2020). Lack of standardised protocols of data collection, storage, and analysis as well as data verification and validation can hinder utilisation and the effectiveness of CGD (Ottinger, 2010).
- **Data management** means making CGD findable, accessible, interoperable, and reusable (Fritz et al., 2019). This involves challenges related to the technical infrastructure (e.g. maintenance and upkeep, limited access to storage for archiving, scalability) as well as to data integration, interoperability (ensuring that CGD can be easily integrated with official data sources and other types of environmental data), compliance with legal and regulatory requirements, and data privacy and security. These challenges require a holistic perspective that combines both technical and legal expertise.
- **Data ownership** in CGD tends to be diffuse, due to the collective nature and the frequent anonymity in data collection. The awareness often remains within the community of knowers. Indigenous knowledge is a case in point (Walker et al., 2021); the specificities of indigenous knowledge production may also question the fitness of standardised ways of eliciting consent for data handling (author's fieldwork, 2020). This diffuse ownership contrasts with the credentialed ways of producing knowledge, such as in academia or the consultancy industry. Data sharing licenses might pose a challenge, unless they account for collective ownership.
- **Sustaining engagement and participation** points to three orders of problems. CGD are typically under-resourced and based on volunteer work. Incentives are needed to sustain participation and motivation over time, making sure people contribute data regularly. Ensuring diverse and inclusive participation across demographics (inclusivity) is also a challenge, as the engagement of CGD is tied to education and access to resources. Finally, capacity building is a key ingredient: expertise, which is often individually owned, can be shared within the community.
- **Vulnerability of CGD makers**. Marginalised knowers, vulnerable individuals and communities, or people in emergency situations, including aid workers, tend to be at the forefront of CGD production (Gutierrez, 2018), but might not find themselves in the ideal conditions to supply consistent flows of high-quality data. Because 'marginalised people are often marginalised as knowers

as well', recognition and participation of citizens in data making are not sufficient conditions to support CGD (Ottinger, 2024, p.200). A data justice perspective can help bring power imbalances and the rights of vulnerable groups to the forefront.

- **Governance and policy integration**. Despite growing in popularity, CGD often lack recognition in policy and scientific circles, making effective use in policymaking uncertain (see, among others, Schade et al., 2021; Berti Suman et al., 2023). Fostering collaboration between citizens, nongovernmental organisations, government agencies, and other stakeholders requires further experimentation. Establishing regulatory frameworks to support and govern CGD in environmental governance is a key challenge, as explored next.

6.3.4 CGD: Towards an enabling ecosystem. Necessary measures to support CGD

Amidst the significant shifts in power dynamics within the 'science-society-policy interface' (Schade et al., 2021), it is crucial to explore how to empower stakeholders in CGD and enhance their effectiveness for environmental and climate governance. This section draws from the frameworks of environmental democracy, data justice, and epistemic (in)justice (Fricker, 2007) to outline the key components of an enabling ecosystem for CGD. It centres 'careful knowing' – an approach inspired by the feminist notion of care, responsive to the needs of marginalised knowers, and complementary to participation and recognition in environmental justice. At the very least, careful knowing means 'developing epistemic resources and standards of evidence suitable to the specific circumstances of frontline communities, as well as the need to care for community members' status and self-regard as knowers' (Ottinger, 2024, p.215).

Starting from a system thinking perspective, it is essential to recognise that data collection and data use are interdependent and cannot be separated. Thus, **an enabling ecosystem mediates between the creation of CGD ('making') and its application ('exploiting') in policymaking and awareness raising**. Here, the concept of an 'enabling ecosystem' goes beyond legislative frameworks (as further explained in Chapters 5 and 8 of this report) and encompasses issues of legitimacy, accountability, and empowerment. In line with system thinking, it attributes responsibility not only to institutional actors (e.g. policymakers, funding bodies), but also to the CGD community itself. By addressing these aspects, we can create an environment that supports effective CGD use in environmental and climate governance.

Six elements have been identified as part of the enabling ecosystem: at the community level, literacy and accountability; at the material level, resources and infrastructure, and at the institutional level, legitimacy and recognition on equal footing.

The first class of elements of an enabling ecosystem for CGD intervenes at the community level.

- **Data literacy as flywheel of participation**. Within CGD initiatives, experiential knowledge becomes meaningful when combined with data skills (e.g., Sander, 2020). The challenge is to 'go from something that's more of an expert culture [data handling] to transferring this to people' (DATACTIVE interview, 2019). Whilst fostering data literacy programmes would also be the task of governments and funding bodies, it rests also on the CGD community to

promote grassroots training, knowledge sharing, and mutual learning.

- **Accountability** within an enabling ecosystem for CGD refers to the mechanisms, practices, and policies that ensure all participants – citizens, researchers, policymakers – are responsible for their actions and contributions to the data collection and usage process. This would involve (a) transparency in data collection and usage, and open communication with CGD participants and data subjects; (b) ethical conduct, ensuring data are collected and used responsibly without causing harm; (c) data stewardship by those managing the data, who must ensure accuracy, privacy, and security; and (d) equitable opportunities for all, particularly marginalised knowers, to participate and influence the data processes. Commonly agreed standards and guidelines would contribute to mainstream accountability within CGD communities and initiatives. Registered Data Altruism Organisations (RDAOs), defined by the Data Governance Act as non-profit trusted entities that make relevant data available within a system of system of safeguards that protect the rights and interests of citizens and companies, have a role to play in driving this process.

At the material level, an enabling ecosystem for CGD ensures that citizens have access to the necessary tools, resources, and support systems to effectively generate, manage, and utilise data. Public funding and provisions facilitating private support (e.g. tax breaks or monetary incentives) are instrumental.

- **Financial resources** include sufficient funding to support the purchase of equipment, maintenance, and operational costs of CGD projects, and to fund training for CGD knowers. Availability of grants and sponsorships from governments, nongovernmental organisations, or private entities would contribute to kick-start projects but also to sustain long-term CGD initiatives and ensure that data remain available and maintained over time. States could retrieve the necessary monetary resources to support CGD capacity-building and initiatives by taxing revenues of data service streams (e.g. data centres).
- **Technological infrastructure** is a sine-qua-non condition of existence and survival for most CGD initiatives. This would include the availability of reliable devices and sensors that citizens can use to collect data (e.g. air quality monitors, water testing kits), access to high-speed internet and digital platforms that enable the transmission and sharing of data, and to a robust system for storing large volumes of data and processing capabilities to analyse the data collected. In this context, it would be important to consider the new data access rights provided by the Data Act to users of connected devices and how this can affect data collection practices in the context of CGD (see more about the opportunities and limitations of business to consumer (B2C) data sharing in Chapter 4 of this report).

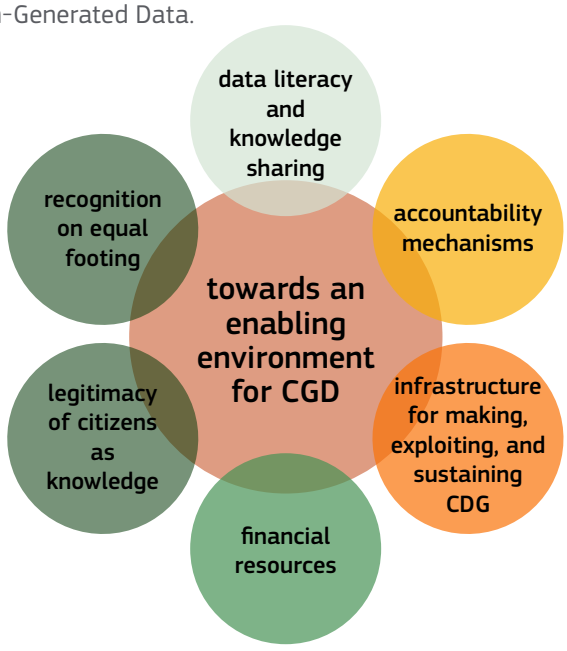
It has been observed that 'the degree and kind of empowerment environmental surveillance supports is determined by the manner in which surveillance data is made meaningful' (Ottinger, 2010, p.221, emphasis added). Thus, the third class of elements for an enabling ecosystem requires **pushing the boundaries of the regulatory framework** to make CGD meaningful.

- **Recognition on an equal footing**. Institutions often downplay the role of citizens and laypersons as knowledge creators. However, insights from multistakeholder governance, where all stakeholders are entitled to participate in policy decisions on an equal footing (cf., Raymond and DeNardis, 2015), can help attribute to CGD the role such date deserves. While disparities in resources and expertise can significantly undermine these ambitions, the institutional set-up would be geared towards attributing CGD the voice it deserves. This recognition would be tied to the development of adequate regulatory standards as a 'boundary-bridging' tool (Ottinger, 2010) for CGD to be recognised as valid sources of information.
- Within CGD, the **legitimacy of citizens as knowers** is fragile and contingent, drawing from various sources such as representation, experience, and even moral authority. It cannot be established in advance or permanently. However, a robust regulatory framework would significantly help in legitimising non-credentialed forms of knowledge. This aligns with the concept of 'supplementary democracy', which supports enhancing existing mechanisms to promote broader civic participation and greater accountability (van Rooy, 2004, p.137). Establishing the 'right to contribute information and have that information considered by appointed institutions' (Berti Suman et al., 2023, p.4) would further legitimise CGD.

Figure 5 depicts the six elements of an enabling ecosystem in relation to roles and responsibilities within the GDDS. It is to be read clockwise. Data literacy and knowledge sharing, in orange, are at the root of CGD, and are the joint responsibility of the CGD community and institutional actors. Accountability (in yellow) rests solely with CGD communities, while technological infrastructure, financial resources, legitimacy, and recognition on equal footing (in shades of green to indicate their incremental effect on CGD initiatives) are primarily the duty of institutional actors.

FIGURE 5.

Enabling ecosystem for Citizen-Generated Data.



Source: own elaboration.

6.4 Conclusions and policy recommendations

Citizen-generated data are non-traditional data sources that are created or collected, analysed, reused, and/or disseminated by individuals and groups outside the remit of the State and the industry, with or without the collaboration of specialists such as scientists and either not-for-profit or in combination with for-profit activities. They are at the core of a demand-driven approach to data production within the GDDS. The lifecycle of CGD includes six stages, divided in 'making' and 'exploiting' data. Citizens as knowers can either collect data that has been passively or actively generated or produce data from scratch; data, in both analogue and digital format, is then analysed. Once generated and analysed, citizens can exploit CGD in three main ways. They can promote environmental awareness and a just, sustainable transition by utilising data for informing policymaking, enforcing law or to raise awareness ('deploying data'). They can use data for co-designing legislation or to develop solutions, including products ('accelerating data'). Finally, they can use data for self-empowerment or to promote virtuous behaviour by others ('activating data'). These stages identify potential transformations implemented by means of data but are neither required nor mutually exclusive.

CGD encounter six main challenges to effectively contribute to environmental governance: data quality, data management, data ownership, sustaining participation, vulnerability of CGD communities, and governance and policy integration. An enabling ecosystem for CGD implementing the principle of careful knowing, centring the needs of marginalised knowers, empowering them and caring for them, would address these challenges by implementing community measures (literacy and responsibility), mobilising resources and infrastructure to support the grassroots generation and mobilisation of knowledge, and adapting the regulatory framework to meet the needs of CGD knowers.

As a result of the analysis, we can identify **three policy recommendations to foster an enabling ecosystem for CGD inspired to careful knowing and data justice**. The objective is to integrate, enable and give legitimacy to CGD, exploiting their potential in the promotion of environmental awareness and a just, sustainable transition.

i. Craft an incentive scheme to foster, support and sustain CGD initiatives redirecting revenues from the data economy (e.g. data centres, tech companies) to grassroots initiatives, including small and medium-sized enterprises (SMEs). This incentive scheme should consider both financial resources and technological infrastructure.

ii. In cooperation with relevant stakeholders, including RDAOs, develop adequate regulatory standards for CGD to be recognised as legitimate sources of information in decision-making processes. Revise these standards on a five-year basis to make sure they reflect the evolution of technology as well as policy problems.

iii. Promote CGD awareness regarding their potential as well as the role of RDAOs by means of EU-wide information campaigns and educational resources, in collaboration with public institutions (e.g. public libraries, the school system) and private entities (e.g. technology companies, the media system).

7

IMPLEMENTING CARE PRINCIPLES FOR ENVIRONMENTAL DATA SHARING WITH INDIGENOUS PEOPLES AND LOCAL COMMUNITIES

7.1 Introduction

Citizen generated data (CGD) on the environment from Indigenous[34] peoples and local communities (IPLC) offer great opportunities for increasing the availability of environmental data as well as hold potential for diversifying the sources of relevant knowledge (Tengö et al. 2021). Following the overview and insights provided by Chapter 6 on 'Citizen-generated data within the Green Deal data space', this chapter delves deeper into one of the most crucial aspects of CGD: the importance of fair and inclusive governance models.

It focuses on environmental CGD that come from communities who have long-term connections and relations with their environments, such as intergenerationally transmitted planting cycles and hunting habits, but also a deep understanding of the soil condition, purity of waterways, migratory patterns of birds, etc. Therefore, environmental CGD discussed here can be classified as a part of Indigenous and local knowledge systems, which also include data academically classified as scientific data, referring to data that have been gathered using methodologies of natural sciences and conforming to their standards of presentation.

IPLC can also gain from generating and sharing environmental data. As an example of a policy implication that highlights the importance of data, Walter and Carroll (2021: 15) ask which data are utilised as evidence for deciding policy priorities. This further relates to Indigenous peoples' will to govern which data are collected from them by outside actors, such as scientific institutions or State actors, and how it is represented. Chiefly considering our scope, it shows the incentives that IPLC might have for contributing environmental data. Contributing to environmental data could potentially help to address environmental concerns recognised by IPLC, as Berti Suman (2021) has conceptualised as a right to contribute environmental information from a broad understanding of the Aarhus Convention of 1998.

However, communities may encounter various risks if they engage in environmental data sharing. Lämmerhirt et al. (2018: 14) observe **the importance of considering data governance around CGD**, since there are several issues of concern around sharing CGD. These include failures to acknowledge data authorship, ownership of data and data infrastructure, and privacy protection. CGD can also pose risks for groups of people, since data might also contribute to increased vulnerability of groups (Daly et al. 2019; Taylor et al, 2017). Similarly, attention has been drawn to conflicts that Indigenous communities experience between protecting their interests in scientific data generated from their lands, waters, and people while supporting, or being subject to, open data and data sharing initiatives due to worries about secondary use of data, issues with bias and social inequity, and limited opportunities for benefit-sharing (Carroll et al 2021; Daly et al. 2019).

Increasing attention to data governance models around CGD offers ways to mitigate these harms and distribute benefits from data sharing to communities. A particularly innovative approach is offered by Indigenous Data Sovereignty (IDS) and CARE principles (Collective benefits, Authority to control, Responsibility, and Ethics) that broadens the scope of issues usually associated with

34. Indigenous is capitalised as a sign of respect and recognition for Indigenous identities and institutions (similar to capitalisation of English, Spanish, etc.).

CGD. They situate IPLC control of data and emphasise the importance of distributing benefits gained from data with communities.

Considering these critiques, the purpose of this chapter is to **identify potential good practices regarding the sharing and use of environmental data generated by IPLC**. I ask the following questions:

1. What are the benefits of including CGD from IPLC in environmental policymaking?
2. What are the main obstacles faced by the institutions and IPLC when utilising and sharing environmental CGD, and what ethical concerns must be considered regarding CGD management practices?
3. How can existing approaches to data governance, such as Indigenous Data Sovereignty and CARE principles overcome these obstacles and what challenges remain?

The next section briefly introduces the scientific discussion and observable benefits from including environmental CGD by IPLC to climate research and policymaking on an institutional level. Section 3 introduces the overarching concept of Indigenous Data Sovereignty (IDS), that includes the data management approach Indigenous Data Governance, and, most importantly considering the aims of this chapter, CARE principles, the guidelines developed for achieving IDS in projects and on an institutional level. Based on expert interviews, Section 4 presents an analysis of how CARE principles were manifested in the cases chosen for this chapter: Local Context initiative, Local Indicators of Climate Change Impact project, Water Data Collaborative, and the MONEC project. After the discussion, the final section will draw conclusions on the research questions arguing that a trust barrier exists between institutions and IPLC that hinders the inclusion processes and utilisation of CGD. While CARE principles are a good starting point for overcoming this major obstacle, the European Commission needs to vigorously address data management issues to facilitate CGD inclusion in policymaking and climate research. This need is becoming ever more acute with the development of GDDS.

7.2 The contribution of citizen-generated environmental data from Indigenous peoples and local communities for environmental information

The inclusion of data from IPLC holds a significant potential for increasing available data from places that are not well represented in the official data sets employed by the public institutions (Bonney, 2021; Reyes-García et al. 2022a, Tengö et al. 2021), thus making them highly relevant for the GDDS. Moreover, with the pressure from climate change, Indigenous and local knowledge is increasingly observed as a part of the solution to understand the local impacts of climate change that complement the statistics on the large-scale changes (IPCC 2022; Reyes-García et al. 2022a). Similarly, IPLC's knowledge contributes to maintaining biodiversity (Reyes-García et al. 2022b). This type of data offers tangible benefits for better-informed environmental decision-making.

Furthermore, the inclusion of data from IPLC has the potential to increase visibility and representation of these groups with local decision-makers and enable their

participation in environmental governance (Bonney 2021, Tengö et al. 2021). IPLC have successfully engaged in community-based citizen science projects to highlight and make visible environmental issues otherwise ignored by the wider society (Berti Suman et al. 2023), highlighting how CGD, and the official acceptance of such data, can create visibility and representation for vulnerable communities. This enables the representation and inclusion of groups that often suffer from marginalisation in environmental decision-making (Massarella et al. 2022).

Institutional utilisation of CGD enables collecting data otherwise difficult to collect. For fostering democratic inclusivity, the enabling of GGD would generate public value through better-informed decision-making.

7.3 Indigenous data sovereignty, governance and CARE principles

To begin the discussion around data governance in the context of CGD, 'data governance' needs to be defined. Data governance in this context refers to the distribution of power among actors over decisions that are taken about data. Micheli et al. (2020: 3) define data governance from the perspective of social sciences as 'the power relations between all the actors affected by, or having an effect on, the way data is accessed, controlled, shared and used, the various socio-technical arrangements set in place to generate value from data, and how such value is redistributed between actors'. Data governance outlines who has the capacity to decide how data 'is accessed, controlled, used and benefited from' (ibid.).

Observing the constellation of power among actors in data governance helps to situate citizens and communities as actors in data governance. Here I focus especially on the emerging forms of data governance that facilitate the participation of communities and citizens in data governance in the context of CGD. The literature on data governance has advocated for the importance of heterogeneous approaches to data governance that go beyond simple public open data or a situation where data are controlled exclusively by a company owning a platform, where data are collected (Carballa Smichowski 2019). The rest of this section introduces the overarching concept of Indigenous Data Sovereignty, the data management approach Indigenous Data Governance. Their benefits are discussed in consideration of how they address the problematic nature of collective data rights. The section ends with the introduction of CARE principles, data management guidelines developed to advance IDS.

7.3.1 Indigenous data sovereignty and governance

Risks regarding the increased use of CGD for policymaking have been addressed in the publication by the European Commission from the institutional point of view (2018). Nevertheless, little attention has been paid to possible risks that IPLC encounter as a result of the increased interest in data that they can offer.

Indigenous Data Governance provides an important contribution to data governance approaches, as it addresses the cultural context of data and emphasises the power dynamics behind the current data acquisition processes (Carroll, Rodriguez-Lonebear, and Martinez (2019). Indigenous Data Governance situates Indigenous peoples'

control of data in the centre. This idea is epitomised through the concept of Indigenous Data Sovereignty (IDS), which Carroll, Kukutai, et al. (2019) define as 'Indigenous people's rights to control data from and about their communities and lands, articulating both individual and collective rights to data access and to privacy' (300). Working towards more inclusive and equitable data governance and IDS requires paying attention to power dynamics that have disempowered Indigenous peoples from regulating what type of Indigenous data are collected, how they are represented, utilised, or shared (Carroll, Rodriguez-Lonebear, and Martinez 2019; Carroll, Kukutai, et al., 2019). It also questions the positivist and universalising approaches to knowledge procurement and propagation and emphasises the cultural context of data. These insights are relevant for all CGD from IPLC because they help to develop data governance that avoids misrepresentation of data, as well as empowering communities to have authority over their data.

In this context, **the new data access rights provided by the Data Act (European Union, 2023) could be considered as a potential tool to empower these communities**, by increasing their capacity to access data generated by their use of connected devices and services. The Regulation 'ensures that users of a connected product or related service in the Union can access, in a timely manner, the data generated by the use of that connected product or related service and that those users can use the data, including by sharing them with third parties of their choice' (Data Act, Recital 5). Nonetheless, exercising these rights might not be an easy task, and the practical impact of the Regulation will depend on the complexity of its implementation and the willingness of both citizens and device manufacturers.

With the increased attention to Indigenous and local knowledge as one of the possible sources of environmental data, **it is fundamental to assess risks that data sharing can pose for these communities** and consider the role of the communities as active participants in the data ecosystem. The problems related to such practices, such as misrepresentation and lack of control over sensitive data, and questions concerning intellectual ownership over traditional ecological knowledge have become more urgent.

Indigenous and local knowledge can have vastly different epistemic foundations, which frequently can lead to misunderstanding and misrepresentation. Indigenous knowledge is 'a complete knowledge system with its own languages, with its own concepts of epistemology, philosophy, and scientific and logical validity' (Battiste, 2008: 500). A debate on epistemological and practical divergences between scientific ecological knowledge and Indigenous knowledge is beyond the scope of this chapter, but it is important to acknowledge the cultural background of environmental data to understand the risk of misrepresenting Indigenous and local environmental knowledge.

Moreover, it is important to recognise the cultural value that Indigenous and local knowledge has for communities. Collected environmental data can also hold religious, spiritual, and social value for individuals and the collective (Robinson et al 2021; Williamson, Provost, & Price 2022). Indigenous peoples have suffered from appropriation of their environmental knowledge for unauthorised and uncompensated commercial purposes through the acts of biopiracy (Robinson 2010). In addition, IPLC have been stigmatised for their socio-ecological practices and have been subject to misinformed

institutional interventions prohibiting their traditional forms of environmental management practices (Ampumuza 2022; Lightfoot et al. 2013). In the current context of environmental data sharing and open data environments, Reyes-García et al. (2022a) identify risks for incorporating Indigenous knowledge in the context of data sharing and open data environments. There is a risk of producing 'extractive' knowledge practices that are characterised by highly unequal power relations between actors, and in which communities are treated as data-producers without any substantial benefit from such participation. Reyes-García et al. state that community-based environmental monitoring initiatives must recognise IDS to prevent the continuation of extractive knowledge practices, which frequently result in the misuse or misappropriation of Indigenous knowledge (2022a: 716).

7.3.2 Collective data rights and non-personal data

Kukutai et al. 2020 highlight the need for **collective data and privacy rights in addition to personal data and privacy rights**. The current focus on the protection of personal data rights fails to adequately address the needs that Indigenous peoples have for data governance, and they call 'for the development and implementation of collective Indigenous privacy laws, regulations and standards' (2020: 655).

Indigenous Data Governance brings into view how the current focus on personal data rights can fail to respond to the needs that communities have for data governance. Similarly, the protection of personal data rights does not account for how sharing personal data has implications for groups of people, as data are utilised with data analytics and artificial intelligence, which focuses on identifying patterns in data (Taylor et al 2017). In this situation, data about an individual affects a broader community of people. Also, Ho and Chuang (2019) argue that anonymisation and individual consent as legal remedies for advancing data rights from the perspective of individual autonomy are currently insufficient tools of protection of data rights.

Another important insight from IDS is how it is used to claim rights to have data for governance as well as governance of data (Walter and Carroll 2021). Walter and Carroll point out the need that Indigenous peoples have for having access to data that could be used for governance by IPLC. This claim is equally important for several communities that could benefit from having access to non-personal data meaningful for a community. For example, this approach could be equally applied for claiming access to data that could strengthen environmental management in the case of agriculture. Small agricultural actors also often suffer from data scarcity due to the fragmentation of data sets in the agricultural sector and inability to access relevant data (Finck and Mueller 2023). IPLC and other communities lack the possibility generate benefits from non-personal data because: 'The paradigm of creation, collection and use is even worse in the context of non-personal data, where there is yet no significant legislation that recognises the interests of communities over non-personal data' ((Nanda et al. 2023: 52). The public value of non-personal data is difficult to realise within the current legislative context.

These notions raise the requirement to address data rights over non-personal data in addition to personal data. Another large issue is the importance of development of collective data

governance and a legal framework to regulate collective rights over data. Data governance models that focus on the collective rather than the individual can also have an additional value by increasing the bargaining power of people in negotiating their data rights (Ada Lovelace Institute 2021: 38). Thus, it would be advisable to facilitate the participation of data-beneficiaries as a collective entity.

7.3.3 CARE principles

An important part of Indigenous Data Governance has been the guidelines developed by Indigenous data right organisations that work towards strengthening Indigenous data rights. The International Indigenous Data Sovereignty Interest Group released the CARE Principles for Indigenous Data Governance (Research Data Alliance International Indigenous Data Sovereignty 2019). The principles consist of **Collective Benefit, Authority to Control, Responsibility, and Ethics**. 'Collective Benefit' calls for data ecosystems that enable Indigenous peoples to receive benefits from the data. 'Authority to Control' signifies that the rights and interests of Indigenous peoples over their data must be acknowledged and the authority to control their data should be facilitated. This also requires that Indigenous peoples and their governing bodies have the right to decide how Indigenous data are represented and identified within data. 'Responsibility' sets a requirement for people working with Indigenous data to share transparently their efforts to utilise data in a manner that strengthens Indigenous peoples' self-determination and for the collective benefit. This includes building up the data capabilities of communities and strengthening Indigenous cultures and languages in the context of data practices. Finally, 'Ethics' brings to the fore that 'Indigenous peoples' rights and wellbeing should be the primary concern at all stages of the data lifecycle and across the data ecosystem' (Research Data Alliance International Indigenous Data Sovereignty 2019). This requires work to maximise benefits and minimise harms caused by data practices, as agreed in the collaboration with Indigenous peoples (Carroll et al. 2020).

Carroll et al. (2021) state that CARE principles are conceived to complement FAIR (Findable, Accessible, Interoperable, Reusable) principles, as they encourage fair participation, outcomes, and attribution of data access, usage, and reuse in a modern data landscape. FAIR principles support data-centric criteria that encourage increasing data sharing amongst entities while ignoring relationships between actors and local environment, power imbalances, and the historical circumstances surrounding the acquisition of data. They are in line with the global movement towards open research and open data. Implementing CARE principles together with FAIR could help to attend to the aforelisted deficiencies of FAIR, as becomes apparent in the analysis of cases in the following section.

7.4 Cases

This section explores four example cases that exemplify community-centric data sharing practices: a Local Context initiative, the LICCI project, the Water Data Collaborative, and MONEC. These cases were chosen, as they epitomise participatory data practices between IPLC and institutions combining CGD, and scientific data produced in collaboration between researchers and IPLC. These cases are exploratory. The focus is on a qualitative study of what the possibilities for data-sharing are. The selected cases exemplify distinct parts of the data ecosystem, and they range from data collection to data sharing. This

qualitative approach was selected to illustrate how participation and the collaboration with IPLC is realised across the data ecosystem. Examples also include cases from different geographical locations and collaboration between researchers and communities and non-governmental actors and communities.

The cases are examined based on CARE principles to indicate community-participation and ethics related to data sharing practices. CARE principles are not a ready-made checklist, rather they are high-level Indigenous Data Governance principles that represent values and ethical principles that aim to guide the work with IPLC (Carroll et al. 2020). For a quick recap, in their essence CARE principles advocate for:

1. **Collective Benefit: How does the project generate benefits for a community collecting data?**
2. **Authority to Control: Is a community able to decide who has access to data and how data are used?**
3. **Responsibility: How does the project build capacity around data practices in a community?**
4. **Ethics: How is the well-being of the community fostered across all stages of the data lifecycle?**

These cases are also complemented by four interviews with five experts on data collection and governance in IPLC settings. The interviewees were Ramin Soleymani and Adrien Tofighi Niaki, who work with the LICCI project team; Robbie O'Donnell, who works for a non-profit environmental technology firm Commons, which forms a part of the steering committee for Water Data Collaborative; Kari Anne Bråthen (UiT The Arctic University of Norway) from MONEC project; and one auxiliary interview from Amrita Nanda, who works at the Aapti Institute based in India.

7.4.1 Local contexts

Local Contexts (2023) is an initiative that aims to facilitate Indigenous governance over their cultural heritage and Indigenous data. They have created Traditional Knowledge (TK) and Biocultural (BC) Labels that 'offer Indigenous communities a tool to add cultural and historical context and cultural authority to cultural heritage content in their own local digital heritage archives as well as in digital archives, libraries, museums and other digital repositories globally' (Local Contexts 2023). Labels add contextual metadata such as community names as well as protocols and permissions that govern the possible use and access of data. In addition to Labels, Local Contexts also created TK and BC Notices that can be added by institutions and researchers, whereas Labels can be added directly by IPLC. Local Contexts provides APIs (Application Program Interfaces) to integrate Labels and Notices to existing data collections management systems.

Indigenous communities can customise Labels to fit their specific cultural context. Local Contexts provide an array of Labels that help communities to claim control and set guidelines for the appropriate use of data. As mentioned, Provenance Labels add metadata on community names, but Labels can also be used to specify the subgroup such as specific family, clan, or collective ownership. Another possibility is to use Protocol Labels that inform data users about traditional guidelines governing the access to data and invite people to respect those rules such as restrictions over gender or season. Moreover, TK Labels include Permission Labels that specify what data-related activities have been approved by a community or if communities are open to collaboration. The latter signifies that the permit for the proposed activity must be directly negotiated with the community, and

it is not agreed on beforehand. For example, Permission Labels advise if data can be used only for non-commercial purposes or only for community use.

Local Contexts is intended to create tools for IDS, and it is planned to be compatible with CARE principles. In that sense, Local Contexts' Labels and Notices provide other actors with tools to apply CARE principles in practice. As a collective benefit, Labels and Notices provide tools for communities to claim their knowledge in digital environments. Labels enforce Authority to Control, as they support IDS and help to protect Indigenous intellectual property. Moreover, they also address the issue of how to present Indigenous data without losing its connection to wider framework of Indigenous knowledge. Labels and Notices help to embed digitally circulating data into the wider framework of Indigenous knowledge by providing context and facilitating relations to communities generating data. They embed Responsibility as a standard research practice, as they envisage the rights of IPLC over their data and provide tools for deeper engagement between IPLC and other actors such as researchers. Ethics are in the centre of Local Contexts that was designed to support IDS and rights of IPLC.

7.4.2 Local Indicators of Climate Change Impacts (LICCI) project + LICCION + OpenTEK (complemented by an interview with Ramin Soleymani and Adrien Tofighi Niaki)

The Local Indicators of Climate Change Impacts project (LICCI) started in 2018, with a five-year funding from ERC (European Research Council), with an aim to elucidate the potential role that local knowledge systems can offer to climate research by:

> ***1.*** *providing data on local perceptions of climate change impacts on climatic (e.g., rainfall patterns change), physical (e.g., shrinking glaciers), biological (e.g., phenological changes), and socioeconomic systems (e.g., crop failure due to rainfall patterns change), and*
> ***2.*** *testing hypotheses on the global spatial, socioeconomic and demographic distribution of local climate change impacts indicators* (LICCI 2023).

The original approach of LICCI 'was classically academic, anthropologic, collecting data on local indicators of climate change impacts from local and indigenous communities across the world' (Soleymani). The data were collected by the collaborating researchers, who worked with indigenous and local communities to gather data. The LICCI project also decided to create a citizen science platform called OpenTEK that enabled anyone to collect and contribute data on local indicators of climate change in their locality.

> *Basically, after the half of the project, we also decided to modify the citizen science platform to allow local communities to collect data that they see more fit for them to help collect data relevant for them to show for policymakers or to plan adaptation strategies or so.* (Soleymani)

So, the OpenTek platform built by the project team was created using Oblo, a free and open-source technology developed by the Institute of Environmental Science and Technology (ICTA-UAB) in the context of the project. The original research project followed the FAIR principles and the basic ethical guidelines regarding data management according to ethical clearance of the Universitat Autònoma de Barcelona and ERC. As the LICCI project gained experience from

their initial work, they realised the importance of IDS and they decided to implement its principles and add TK and BC Notices in the OpenTEK platform. Furthermore, it became apparent that there was a need for a methodological change on what type of data are collected to increase the value and benefits that IPLC could gain from data.

> *OpenTEK platform is using the similar methodology as LICCI researchers were using when they were doing fieldwork. This implies Western scientific approach to data collection and to understanding … to looking at impacts. So, to give an example, a lot of questions were framed around impact on systems, impact on biological systems, or physical systems or atmospheric systems.* (Soleymani)

To facilitate local participation, the project team launched a new proof-of-concept project, LICCION, that was focused on the impact on local livelihoods and based more on IPLC's needs and realities. LICCION developed community-centred data platforms with CARE principles in mind based on the same software, Oblo, that are now in beta-testing. LICCION website states that: 'The purpose of these new domains and surveys is to enable customary and community-led research and evidence-building on climate change while upholding Indigenous data sovereignty principles and values'.

> *For Oblo platforms, we turned that upside down, so these custom-made surveys and custom-made platforms, we started instead with impact on livelihoods, because that was of the most concern for individuals. In the end of day, many communities were not interested in just to record what is happening in river near-by, they were more interested in saying but our impacts are that the houses are being flooded in yearly basis. It was about reframing the surveys, so they were more relevant for their livelihoods.* (Soleymani)

As a result of the experience of working with data from IPLC and the same time dealing with the data standards set for academic publishing and large centralised data repositories, the project started developing a toolkit that could advance the collaboration with IPLC.

> *So, this was two-year proof of concept project and already during that project, when we worked with three organisations from Southeast Asia, we already thought and discussed Indigenous Data Sovereignty with different organisations. But we didn't have time to bring that in. Because projects are always tight in time. Because of that, we applied for another proof of concept, which we do now, where we proposed to develop a toolkit. Toolkit that could help to bridge problems between implementing indigenous data sovereignty and large centralised data repositories that we have everywhere also in Europe and that are not very fit to bring these modifications and concepts.* (Soleymani)

To respond to this problem, they started to develop a toolkit 'that contributes to better data governance and upholds principles of IDS in the context of open-data research'. The adding of BC and TK Labels and Notices was further simplified by the project team by creating a simple website, where the data managerscould link labels from the Local Contexts hub (a platform built by the Local Contexts for local communities) and applying them to an open-source repository system for scientific data publishing,

Dataverse. To increase the control that IPLC have over their data, the files remain private with only metadata directly accessible by the researching party, who, when finding something of interest, can then go through an application process to receive it. This strengthens the Authority to Control, as the IPLC can readily share data in full knowledge that they retain control over who and for what the data are thereafter used for.

> *We are showcasing the toolkit and show-off a little bit on how to share research data with CARE principles. We used LICCI-dataset as an example. We demonstrated with LICCI-dataset. A lot of abstract data, high-level data, which doesn't really give out details; they will be FAIR and openly shared, while more detailed data, we try to bring in some parts of Local Contexts Labels and try to give some management rights to communities.* (Soleymani)

The Ethical compliance of the LICCI project is well scrutinised, and Authority to Control has been addressed in OpenTEK with TK and BC Labels from Local Contexts and LICCION, where IPLC have complete control over their data in their own platform. With LICCION the collaborating IPLC have complete control over their data in their own platform. They can access data and they can choose whether to make data public or maintain privacy.

With Collective Benefit and Responsibility, the end-results of the projects are less visible. LICCI did not have integrated requirements for Responsibility, though some researchers in the project created educational materials for IPLC and helped to build networks for communities such as connections with NGOs. OpenTEK only provides the platform, so there is no further capacity building between communities and the project. LICCION, however, builds the capacity in communities to utilise data for their own benefit through the collaborative development of their own data platform.

In theory, the data collected could be used by the communities for their own needs, for example, for influencing regional environmental policy and seeking funding for climate change mitigation and adaptation. At the same, LICCI did not include IPLC in the original research design, which limited the usefulness of data for IPLC. OpenTEK had the same shortcoming, and its methodological choices that followed the LICCI project did not respond to the needs of IPLC, which limited its impact and usability.

> *This is why I am saying in practice vs. theory. How it can benefit them ultimately, that data is no longer for academic research purposes but is specifically for communities.* (Niaki)

7.4.3 Water Data Collaborative (interview with Robbie O'Donnell)

Water Data Collaborative (WDC) is a United States based non-profit focused on water data collection. Founded in 2017, it facilitates water data sharing from citizen science community groups by providing technical tools (such as APIs) and education and advice on data collection and sharing. WDC acts as a mediator between community science groups and institutions such as state governments, federal governments, environmental protection agencies and academic research institutes. In practice, this means providing education to both sides with different focuses. For institutions, they create awareness of the use of citizen-generated water data and teach them how to utilise such data. For community science groups, they provide advice on best practices of data collection and data sharing, technical tools for data sharing and education on data sharing and collection.

The focus of WDC is to facilitate the applicability and accessibility of CGD on water so that this data could be readily utilised by institutions. As remarked by O'Donnell, in the US only 30–35% of waterways are monitored by government institutions, leaving 65% unaccounted for. The filling of this enormous data gap is where CGD by IPLC come to play, and without their input there could be no realistic estimation of the national condition of waterways. One of the major challenges for WDC is simply to help the data-gathering groups to integrate their results via APIs to national platforms where it can be accessed by institutions.

> *Even though there is all this great data, and I think it is the same in the Europe and all the across globe, but there is a bunch of really good organisations collecting all this data in their backyards, so to speak, but they don't know how to get that data where it needs to be, so to speak. It is not living to its full potential.*

The institutional barriers of data-applicability of CGD are also at the very heart of WDC, as institutions and academia have certain reservations about CGD quality. Furthermore, specifically considering traditional knowledge of Indigenous peoples, data can have religious and cultural significance. To overcome such barriers, WDC offers education on water data collection, management, and interpretation. They also educate institutional actors on CGD use to help them recognise the potential of such data. In the future, WDC plans to work on creating certification processes for collaborative groups for establishing quality standards and uniformity of data.

> *If we can build standards, literacy and education in the collection, management, interpretation and all those other things, of environmental data, in our case water quality data, we can start building a bridge between those groups and the policymaking groups.*

Through standardisation and systemisation, the data could be readily used by institutions and academia. Through the application of data standards, the trust-barrier can be surpassed from one side, but the issues of trust need to be taken seriously from the side of IPLC as well. A significant amount of CGD collectors come from marginalised groups that have historic and concurrent reasons to be wary of sharing any information considering their environment that could be used against them by institutional actors.

> *Like tribal groups, some tribal coalitions or organisations don't want to share any of their environmental data with the US government, nor can I see why they would want to do that. They're just afraid that it is gonna be used to … hurt them in a way, which is very attainable considering the history of those type of relations in the US, Canada and abroad.*

CARE principles have a significant role to play in building trust in communities to share their data with institutional actors. With an eye on Collective Benefit, WDC works towards bringing benefits to community groups by increasing the impact of their data on public decision-making and policy. Its goal is to facilitate the sharing of data collected by citizen science groups with institutions.

> *One of the things that we're trying to promote, my organisation the Commons especially, because we're environmental data group, is to try to get these community groups to utilise their data to its fullest extent while also retaining data sovereignty. So, we can get a group to feel comfortable with collecting this environmental data, feel comfortable*

> *using it, feel comfortable interpreting it, and they have full control over it.*

Adhering to Authority to Control, WDC does not compromise the control of data by communities, who remain in the full control over decisions over their data. Citizen science communities retain their data sovereignty. WDC facilitates sharing data through the provision of APIs that enable direct sharing of their data to the data portals that institutions use. WDC also provides education on the importance of data sovereignty for citizen science groups if needed.

> *We're trying to build data collectors to have these skills, but then also feel like that they are not like ... being 'oh we're here only to give decision-makers data', like that's their data, they're controlling it, they're making decisions about it, they're advocating on behalf of it, they still retain autonomy, so to speak.*

With Responsibility and Ethics, the entire process of grassroots collaboration and education on data-collection and utilisation works in tandem to simultaneously ensure the data quality to meet institutional standards, and get the CGD collectors to, not only remain in full control of their collected data, but also to understand how to utilise their data to their benefit. This further strengthens data sovereignty, as true control over collected data has no significance without proper knowledge of data use and value. The application of CARE with FAIR is yielding promising results for WDC.

7.4.4 MONEC – Data collection together with Sami and local communities (interview with Kari Anne Bråthen)

MONEC is an ecological research project that investigated the costs and benefits of managing native invasive species, particularly crowberry, in the context of the socio-ecological system of reindeer husbandry and sheep farming. The project's purpose was to research plant ecology and biodiversity in the Arctic areas in collaboration with reindeer herders and sheep farmers in Norway. The goal was to understand the drivers of environmental change and identify practices supporting local biodiversity. The reindeer herders conducted experiments regarding the management of crowberries, where they could use their knowledge on their land to choose where they would apply the practices. Overall, the research design was modified based on collaborative consultation with IPLC.

The interpretation of the data and representation of findings was also collaboratively discussed with IPLC. This helped to present the data according to the local understanding and needs and resulted in the planning of white paper that focused on providing feedback on the definition of biodiversity in the Norwegian law regarding the reindeer herding/Arctic areas. The current definition of biodiversity is lacking from the perspective of ecologists involved in the project. Also, the drivers of environmental change are based on indicators related to reindeer, and the actions to support the biodiversity are similarly based on interventions on reindeer husbandry. This results in an action that does not necessarily create an impact on the real driver of environmental change. The collaboration with IPLC helped to understand the socio-ecological reality of environmental change in the Arctic area.

IPLC could also help to show their environmental practices and the use of natural resources through data. Reindeer herders have been reluctant to produce data on these practices in the past, because of the feeling

of distrust, as they fear that data would be used as a basis of intervention that does not account for the local needs. For example, the misinterpretation of reindeer herding practices could result in land conflict, as the local authorities would perceive a critical piece of land as unutilised due to the poor understanding of the needs of the reindeer husbandry.

> *And there is a history of being very careful of giving any information about the land or the area from the reindeer husbandry's side, because there is a little trust in the society to give away information about, to say okay you don't need this land, you only use it this and this much, so we can take that land. So, they are incredibly careful, right of how they want to portray. And then that also has a consequence because they cannot say when their core areas for reindeer husbandry are at stake. They don't really have means of saying that you cannot take that area, because that is the most critical area of our all area.*

In this case, local knowledge informed the research data, even if the scientists do the collecting. This resulted in the promising commentary on how to modify the understanding of biodiversity in Norwegian law to better respond to environmental challenges.

From the point of view of CARE principles, the first principle of Collective Benefit was clear in the project. The study had scientific objectives, but it also focused on addressing the problems reindeer herders faced in the context of current environmental change and Norwegian law. This is manifested by the white paper addressing law on biodiversity and reindeer husbandry. It has been planned together with communities, and aims to bring into focus current problems that misunderstand the concept of biodiversity and environmental change, both from the point of view of ecology and reindeer husbandry. This process is ongoing, which means that there is no way to assess the impact, but the project aimed to also address collective benefit for communities.

Authority to Control was not addressed in the project, which followed the principles of open science. Accordingly, data will be made public and open access, once the process for sharing is complete. The research team clearly communicated objectives and research interests from the start to communities clearly and transparently. Communities consented to this course of action and making research data public. Therefore, the principle of Responsibility was followed. Communities were informed about findings throughout the research process and participated in interpreting data and designing experiments. Hence, they could contribute to better representation of their socio-ecological practices through environmental data.

Nevertheless, the research project followed good practices of current scientific ethos of open science, which to some extent is contradictory to the principle of Authority to Control. Bråthen reflected that their research interest on plants contributed to non-conflictual response. In her opinion, the data were more neutral. The contents were not of cultural significance, which would prevent data from being shared further, and communities agreed to having data as open access.

The project also considered the well-being of communities, which is shown by intensive communication and engagement with them as well as the care that was shown in interpreting the results from their point of view. Regarding Responsibility, MONEC invested in collaboration with IPLC, and they included representatives of IPLC in the project that influenced the design

of the project experiments and the representation of data. Environmental data were interpreted also from the point of view of socio-ecological practices of communities to portray an accurate picture of the situation. Risks were mitigated through consultative and collaborative processes.

TABLE 8.

Analysis of CARE principles across the cases analysed.

Project	Collective Benefit	Authority to Control	Responsibility	Ethics
Local Contexts	Provides tools for communities to claim their knowledge in digital environments.	Supports IDS with TK and BC labels that can be applied in data repositories.	Builds possibilities for the deeper engagement with IPLC and other actors such as researchers. Educates also the wider public about Indigenous data rights and interests.	Local Contexts was planned to respond to Indigenous rights and needs over data.
MONEC	Increased understanding of the relationship between biodiversity and socio-ecological practices of IPLC. MONEC plans to release a white paper to influence policy-making.	The DMP is agreed in the beginning of data collection. The DMP follows the FAIR principles. Communities do not have any further control over its use.	MONEC invested in the collaboration with IPLC and they included representatives of IPLC in the project that influenced the design of the project experiments and the representation of data.	MONEC aimed to bring benefits to communities as it promoted the better insitutional understanding of socio-ecological practices of IPLC. They included IPLC in the interpreation of data and writing of publications to avoid misrepresentation.
LICCI	Increase data on local indicators of climate change that can be used by communities for influencing policy-making, advocacy, and seeking funding.	The DMP is agreed in the beginning of data collection. The DMP follows communities for influencing the FAIR principles. policy-making, advocacy and Communities can access seeking funding. data, but they do not have any further control over its use.	No integrated requirements for responsibility. Some researchers in the project created educative materials for communities and helped to build networks.	LICCI fulfills the ethical requirements of Universitat Autònoma de Barcelona and ERC. The primary objectives were aligned with scientific goals that were not defined with IPLC.
OpenTEK	Increase data on local indicators of climate change that can be used by communities for influencing policy-making, advocacy, and seeking funding.	OpenTEK utilizes TK and BC notices to show interests of IPLC on their data. IPLC that contribute their data can decide to make their data public, restricted, or private.	OpenTEK provides only the platform. No capacity-building between communities and the project.	OpenTEK provides a platform to visualize experiences of IPLC to promote justice, but there is no collaboration with IPLC in the development of platform.

Project	Collective Benefit	Authority to Control	Responsibility	Ethics
LICCION	Increase data on local indicators of climate change that can be used by communities for influencing policy-making, advocacy, and seeking funding.	IPLC have complete control over their data in their own platform. They can acess data and they can choose to make it public or keep it private.	LICCION builds the capacity in communities to utilize data for their own benefit through the collaborative development of their own data platform.	Data platforms are developed based on the specific needs of IPLC that are defined through a collaborative process.
Water Data Collaborative	Have more data available on local waterways and increase the integration of citizen science data into the larger official data-sets through the use of APIs. WDC educates institutional actors on the use of CGD to maximise its potential for policy-making.	Participating communities have complete control over their data. WDC provides education on data sovereignity.	WDC increases community data capabilities, as it provides education on water data collection, management and interpretation. It also recommends data tools and best practices.	As WDC promotes the increased use of CGD in water management, they promote environmental justice through better representation of local waterways in the official data-sets. Their work aims to build trust between actors.

Source: own elaboration.

7.5 Discussion

From all expert interviews, the main catch was the issue of the **trust-barrier between IPLC and institutions**. IPLC do not necessarily trust the State institutions to act on their behalf and with their best interest in mind, which prevented them from sharing data. Similarly, State institutions often question the quality of data produced by IPLC. While significant attention had been paid to legitimising CGD as scientifically valid data (see European Commission 2018), it is equally important to build trust that IPLC have for sharing their data. Promoting data sharing practices that abide by the CARE principles is a part of the process of creating trust.

As argued in section 2 of this chapter and presented through case studies above, **environmental CGD from IPLC can positively contribute to a wide array of environmental challenges. CGD can encourage the inclusiveness and representation of marginalised communities** through data in policymaking. Nevertheless, this requires defining the value and Collective Benefit that data potentially have also for communities, because otherwise the project might fail to attract participation from IPLC and perpetuate the experiences of marginalisation for them.

A perceived value and benefit are significant enabling factors for data sharing. There is a gap between a theoretical value from data sharing and current reality. This relates also to the current **lack of institutional uptake of CGD**. Nanda et al. (2023), for example, argue that it is advisable to open the institutional uptake of CGD in order to promote data sharing, as it strengthens the incentive for sharing such data. An increase in the institutional use of CDG would increase the perceived value created by data sharing for communities if they were more represented in environmental data. Defining Collective Benefit and value for IPLC, together with the application of other CARE principles, is therefore crucial

for empowering communities through data practices. This point was well articulated by Amrita Nanda (Aapti institute) during the auxiliary expert interview:

> *There's a sort of vagueness or lack of clarity with which even governmental states are able to define public value from data. I think one of the first steps that needs to happen that we perceive in small ways is to understand how communities can define this value. (...) Being able to articulate this value out of data, because we are speaking of people living in open spaces who do have all sort of bare minimum understanding of what may be the potential of this information and knowledge can be, but not quite that link how it can directly impact their day-to-day lives.*

The analysis of the cases shows that the participation of IPLC from the beginning can contribute to bringing collective benefit. For example, according to the interviews, LICCI and OpenTEK were limited in their possibilities to produce data that would serve the needs of communities in addition to scientific objectives, because they lacked the initial participation of IPLC in defining collective benefit.

In comparison, MONEC (also an academic project) succeeded in producing results that were useful also for IPLC, because it had a strong participatory component throughout the project. Bråthen emphasised the importance of defining what biodiversity means for reindeer herders and going through conversations, where this definition is contrasted with ecological understanding of the concept. This deep collaborative engagement contributed to successful data collection and interpretation since it helped to establish a mutual understanding of biodiversity as the object of study. Similarly, reindeer herders had experienced in the past that their socio-ecological practices of land use were misinterpreted by State institutions, which is why they avoided sharing data. Also in this case, collaborative engagement and interpretation of data related to reindeer herders' use of natural resources were key factors in producing an accurate description of their land use practices. **Collaboration between reindeer herders and scientists helped to create a better understanding of environmental data that established a connection to socio-ecological practices. Chiefly, the collaboration helped to produce high-quality data**.

The second CARE principle, Authority to Control, brings the focus on IDS and on the possibilities of its application in the context of large-scale data sharing, which holds relevance also for the GDDS. This chapter delivers the argument that IPLC have strong incentives to increase their control over their data, because of possible misuse and misinterpretation. From the cases, the problems in implementing IDS are clearly observable in the current context of data management practices, especially where data were gathered with researchers. Neither MONEC nor LICCI had initially any type of plan for upholding IDS. This was due to the lack of knowledge about the concept, as it was not something that researchers were familiar with, and they followed conventional academic practices of data management. Adrien Tofighi Niaki stated that they were only following GDPR (General Data Protection Regulation) protocols, which reflects the lack of awareness around the possible issues around non-personal data that hold value for communities. Niaki also emphasised how this reflects a broader trend in academic settings using the ERC-funded projects as an example:

> *One of the things that we've done as well in this project, is to look at different data management plans of different European Research Council, ERC-funded projects, data-management plans for project that were working with indigenous peoples and local communities. And it would be important to make a note somewhere in the final research results that the European institutions are really behind in terms of ensuring Indigenous Data Sovereignty practices or governance practices. What we find that a lot of data management plans that exist do not align at all with CARE principles and so many projects can just … as long as they abide by the GDPR, they are not … it is the only condition that they have. And in that sense I hope that this research that you're conducting can push a boundary a bit to say … not only with the European Commission, but also with the European research institutions that they also need to be more concerned about data sovereignty and data governance for Indigenous peoples, because there are collective rights that need to be respected, and that means to be more aligned with CARE principles, if possible. So, I am not sure if that is something that at the ERC-level that can be pushed for or if it must be done at the institutional level for each university but is millions of euros of funding going towards projects that fail to consider these rights. We were one of them, I think, and it is really crucial that we change this direction.*

Therefore, the lack of knowledge about IDS and CARE principles forms a barrier to their implementation. This could be addressed by adding IDS and CARE principles as requirements for the ERC funding for the projects that work with IPLC. This could help to build awareness of IDS and CARE, thus increasing their usage.

In comparison, civil society organisations such as WDC and Local Contexts were conscious of data sovereignty and Authority to Control, as were likewise the additional projects developed by LICCI team, LICCION and OpenTEK. WDC's operation model that provided the APIs did not compromise the data sovereignty of communities but provided communities a chance to have more visibility for their data if they so wished. On the other hand, Local Contexts does not focus on generating data; its primary goal is to create tools that will increase IDS in the digital environment. This endeavour is succeeding, as both LICCION and OpenTEK utilise TK and BC Labels and Notices developed by Local Contexts. Overall, the use of TK and BC Labels and Notices provides a promising tool that could be applied also in the GDDS, although **there remain questions such as who would act as an intermediary that gives access to datasets**. Further exploration of the potential role of neutral data intermediaries and data altruism organisations would be needed in the context of IDS.

With respect to **Responsibility**, the primary positive action observed in the cases is the importance of education for supporting data capacities of IPLC. It would be especially beneficial to bring benefits of data for IPLC and other marginalised communities to increase their representation and prevent their further marginalisation in a society that increasingly relies on data. This requires **building capacity around data in communities**. Education in data is crucial. The cases show how giving back to the communities in terms of education in data management and gathering techniques have a twofold positive effect. For one, it strengthens

the scientific legitimacy of CGD, while on the other hand, it franchises the collectors and marginalised groups to better participate in civil politics, environmental decision-making, and democracy in general. From interviews concerning the LICCI, OpenTEK, LICCION as well as WDC, the importance of building capacity around data becomes apparent. Niaki, for example, observed that OpenTEK was restricted, because IPLC do not necessarily have the same possibility to start collecting citizen science data, as they lack proficiency on how to operate the platform.

The problem of obtaining sustainable funding also hinders the application of Responsibility. While a research project can usually obtain a grant for between two and six years, the implementation of CARE principles, foremost Responsibility and Collective Benefit, require a long-term commitment, not only on the part of IPLC, but also from the institutions. At the end of the day, five years is a truly fleeting period for creating and establishing data collection, management and sharing practices with IPLC. The LICCI team, for example, is seeking new grants for continuation of their work. From an institutional perspective, to properly utilise CGD with CARE principles would require a long-term approach and contingency planning for projects and, most importantly, their funding base.

Ethics requires placing the well-being of IPLC at the centre. In this regard, the participation of IPLC is crucial. Overall, it is an ethical commitment to guide other actions, and its realisation depends on the case, since its meaning should be defined by a community itself rather than outside actors. It is also good to point out that not all data are equally sensitive. For instance, MONEC did not implement IDS but rather IPLC, who participated and did not perceive this as a problem. Bråthen speculated that this was due to the topic of concern being plants that were not a politically sensitive subject. This shows how data sharing requires diverse types of solutions depending on the type of data. From an ethical point of view, it is crucial to agree on and communicate the possible benefits that data can bring to IPLC to enable data collection in the first place.

Focusing on citizens and communities as data-providers, we need to encourage inclusiveness and representation through data: **empowering communities through data rather than treating them as passive actors in the process**. As we include CGD from IPLC, we should also consider their overall role in the data ecosystem. How do we bring benefits from data sharing directly to IPLC? Representation through data can help to bring better services. Access to non-personal data that holds significance to IPLC can help their advocacy work on environmental problems. Contributing environmental data can help IPLC become active agents in the environmental governance of local resources and help to bring environmental justice in the case of environmental crimes. Overall, IPLC can use data for their own purposes to govern their own localities better.

7.6 Conclusions and recommendations

As argued throughout this chapter, the benefits from an institutional point of view of including CGD in environmental policymaking are self-evident. Though environmental data are being scientifically collected in ever-increasing amounts, this capacity can only depict a fraction of the overall reality of environmental conditions and climate change. Additionally, to fill in the data gaps, CGD furthermore offers new ways of depiction

and analysis, access to intergenerational cumulated knowledge, and perspectives on viable solutions. The benefits are even more pronounced when observed from the perspectives of democratic processes and meaningful participation in local environmental policy-making, if environmental CGD from IPLC are included in a cooperative manner that adheres to CARE principles. In the case of IPLC, they might have incentives to retain stronger control over their data due to their historical marginalisation. IDS represents an active struggle to overcome the imbalance in the control over data. **To include environmental CGD, there is a need to render citizens as visible actors in the data-ecosystem**. If IPLC are reduced as data-producers, rather than active participants across data-ecosystem, there is a risk of further marginalising vulnerable communities. CARE principles represent data standards which aim to embed IPLC agency in the data ecosystem and encourage their participation.

The previous subchapter identified the two-way trust barrier between institutions and IPLC as the main obstacle for utilising and sharing environmental CGD. Whilst this barrier in its multiple forms – ranging from distrust on data quality to fear of antagonistic exploitation – cannot be overcome with a clear-cut checklist with tick-boxes, inclusion of IDS and CARE as guiding principles can have a marked positive impact on projects that work to include CGD.

This chapter has analysed cases that involve participation of IPLC groups in data production, management, interpretation, and sharing from the perspective of CARE principles. **CARE principles provide a promising approach for increasing trust, incentivising data sharing, and building inclusive data practices**. However, there is a lack of awareness of their existence, which prevents their implementation, especially in the European Union. Implementation of CARE could be encouraged together with FAIR principles by integrating them as requirements for the ERC funding for the projects that deal with IPLC.

This analysis has provided some existing practices that can be implemented to overcome the challenges of data sharing by applying CARE principles:

Firstly, participatory mechanisms, such as consultation with IPLC on the issue of data, help to define collective benefit and value for IPLC. Increasing the institutional reliance on CGD can also help to bring collective benefit from data for communities.

Secondly, technical tools exist, such as TK and BC Labels and Notices that can be utilised also in large-scale data sharing to increase the control that IPLC have over their data. Labels and Notices can be applied to centralised data repositories to show interests and rights of IPLC over their data.

Thirdly, it is important to support education on data capabilities for IPLC to increase their involvement in the data ecosystem and prevent their marginalisation. For example, civil society actors and researchers, who already work with IPLC could be active in this area, if they are provided with funding focused on data education for IPLC.

With the development of the GDDS, this chapter argues strongly for combining CARE and FAIR principles together in the planning and execution processes, for the two-way benefits this would create for both institutions and civil society.

8 LESSONS LEARNED AND POLICY RECOMMENDATIONS

8.1 A Systemic Data Justice interpretation of the findings

In this final chapter, we discuss recommendations and strategies to govern and share data within the context of the common European Green Deal Data Space (GDDS) and the INSPIRE Directive, to drive the digitally empowered transition to a single data economy that addresses sustainability while being socially equitable, trustworthy, and just. As argued in the first chapter of this report, we took a combined approach by adopting systems thinking, as well as placing importance on governing data through a data justice approach.

In this chapter, we introduce what a consolidated approach could look like, which we term **'Systemic Data Justice'**. This approach studies the interdependencies in data ecosystems to examine what kinds of interrelations emerge when these ecosystems are studied as a whole, and not in isolated parts (Meadows 2008). In this approach, issues of equity, accountability, and fair representation are foregrounded to ensure that data and accompanying technical infrastructures do not exacerbate existing social inequalities but rather facilitate public interest and the common good (Dencik 2019, Heeks and Renken 2016, Taylor 2017).

Approaching the GDDS from this vantage point places an emphasis on thinking of relations, interconnections, as well as the contexts within which a fair, just and sustainable transition can emerge. The European Strategy for Data envisages common data infrastructures and governance structures to enable data pooling, access and sharing to boost data use (European Commission 2024). However, in order to achieve this, a Systemic Data Justice approach requires examining firstly, what are the conditions under which data can be exchanged and who are the actors participating in this data exchange; secondly, what are the ways in which to ensure not only high-value and high-quality data, but also high usability of the data; thirdly, what are the governance structures that guarantee access and usage rights and ensure empowerment for people and communities.

A Systemic Data Justice approach introduces the critical importance of studying different kinds of interactions (Meadows 2008) between the economic, social and political consequences of a digitally empowered transition, as well as the role of regulation, incentives, and technical infrastructures to facilitate data sharing. These interactions are not linear, but are dynamic, and complex, and reflect the power dynamics that already exist in the system whether in terms of access, availability, or in the subsequent use of data.

With the revision of the INSPIRE Directive being launched through the GreenData4All initiative, we propose through a Systemic Data Justice approach to study cross insights from the chapters laid out by the authors. These are organised along various aspects addressed by the following questions:

1. **The purpose of the Directive**
   - a. What are the ways in which the INSPIRE Directive needs to be adapted to address the mission of the Green Deal?
2. **The nature and categories of data**
   - a. How should data be categorised in the INSPIRE Directive to move beyond being provider-centric?
   - b. What kinds of understanding of data are currently missing, and how can these be included and represented to facilitate greater data use?

3. **The types of interactions between stakeholders in the system**
   a. What are the ways in which participation between stakeholders could be facilitated?
   b. What governance, legal, social, and technological mechanisms should be established to generate trust both in the data and the processes?
4. **The levels of implementation**
   a. What kinds of governance mechanisms are necessary at a centralised as well as a decentralised level with respect to the role of the European Commission?

These groups of questions will be addressed in the next section in the form of a series of cross-cutting insights extracted from the various inputs by the authors. These insights are designed to form the basis for policy recommendations for the revision of the INSPIRE Directive in light of the current policy context and rapidly evolving technological landscape.

8.2 Cross-cutting insights and policy recommendations

This section outlines and discusses five cross-cutting insights emerging as generalisable reflections across the spectrum of topics covered in the different chapters. While these insights have been discussed throughout the research process and subsequent workshop, we acknowledge that the main contribution of this document reflects primarily the rich reflections and recommendations provided within each of the chapters by the experts[35]. We propose these five recommendations as guidelines for lines of action and overarching principles underpinning the development and establishment of the GDDS from the current INSPIRE Directive.

8.2.1 In the context of the GDDS, INSPIRE would require a shift in purpose from public sector data to public interest data

The INSPIRE Directive, as previously discussed in this report, led to the establishment of an infrastructure for the management of spatial data relevant to environmental policies in the EU, along with a legal framework, as well as technical and governance structures for Member States to implement. By nature, the Directive required a high level of specialisation by data providers from the public sector, leading to a focus on data discoverability, accessibility and interoperability, and less about how data would be used. However, over the years, with the emergence of a variety of data sources beyond the public sector, as well as developments by the private sector in extracting value from spatial and non-spatial data, there is a need to revisit the emphasis on public sector data and to examine the emergence of data ecosystems.

With the mission of the Green Deal to 'leave no person or place behind' there is a need to develop more 'sustainable, flexible, and agile data ecosystems' which can encourage greater openness and participation from non-traditional actors (Kostev et al., 2021). This involves **including novel data sources such as business data or citizen generated data, as well as removing legal,**

35. It is noted that the purpose of this Section is not to repeat the findings and recommendations of each individual chapter.

technical and literacy barriers through the development of agile frameworks that can encourage other actors such as academia, businesses, or citizen groups to participate in data sharing. Introducing a plurality of actors in the different phases of the data lifecycle (production, collection, sharing, using), indicates a shift in emphasis from the public sector to a wider variety of actors. However, from a Systemic Data Justice standpoint, evaluating whether data sharing takes place in the public interest is critical in shaping this ecosystem. The shift proposed is therefore from focusing on public sector data to public interest data, i.e. entailing a public interest in the use of the data itself.

Furthermore, the cross-cutting nature of the Green Deal forces us to think of a related data space drawn upon its key objectives, rather than specific domains of data. The initial questions for many approaching this field revolve around: which data fall within the scope of the GDDS? Is environmental data enough? Should the data be environmentally relevant? If so, how can we define the boundaries of the scope of the GDDS? What does moving from public sector data to public interest data entail in practice? What kinds of actors are able to produce such data? Should it include both spatial and non-spatial data?

When reflecting on these questions, this study argues that the GDDS should be positioned as an ecosystem fostering a data sharing and use economy to achieve the Green Deal objectives that goes beyond an organised repository and associated services and governance mechanisms to promote data sharing. The goal of the GDDS would be not only to ensure that the EU is a competitive digital market that decouples resource consumption from economic growth, but also does so in a manner that takes into account societal inequalities across people and Member States, who experience differentiated impacts in the green transition.

In summary, while the general definition of data spaces applies to the context of the GDDS, its mission and key objectives are different. We argue that the main scope of the GDDS is to foster an inclusive, fair and just data economy to help achieve the Green Deal objectives.

The main outcome of this reflection points towards the scope of the emerging GDDS. From a Systemic Data Justice approach, the shift recommended is from only focusing on collecting and sharing environmental data (e.g. a repository of air quality data) to also actively considering data sharing and reuse to address the Green Deal goals (e.g. promote sustainable transition and resilience). This leads to including a focus on demand-driven data streams which, in turn, can be driven by different elements ranging from emerging environmental monitoring laws, due diligence regulations, to market opportunities. However, this study continues to acknowledge the value of those so-called 'data first' mechanisms as well as the difficulties in fostering a 'mission first' approach instead. Both data-driven experimentations, e.g. in the form of sandboxing or hackathons, and legal initiatives such as the High Value Datasets (HVD) implementing Act as part of the Open Data Directive, provide opportunities for data first innovations through establishing datasets whose publication and sharing should be prioritised based on their innovation/impact potential. However, it can be argued that both of these examples are aligned with integrating a demand-driven perspective in an ecosystem that to-date has been dominated by a supply-driven approach. This demand perspective is integrated by the HVD through prioritising

open data sharing based on use potential, whereas data contests and experimentations are proven mechanisms for driving open data supply from engaging with the open data users (i.e. the demand side) (Kitsios and Kamariotou, 2018). This study proposes a step forward in this discussion, which revolves around placing more emphasis on data use, and beyond open data.

Prior to that, a deeper reflection on 'public interest' should be undertaken. Public interest data remain difficult to capture and manage, since which data fall under public interest is not defined a priori and can assume different meanings based on the context and the stakeholders involved. The overarching question is: **what does public interest data mean?** 'Public interest' is a polycentric and dynamic concept contingent on several factors. It is flexible and a potentially contested notion that can be understood in terms of achieving values of the common good, shared societal interests and promoting greater utility for members of the public through democratic deliberations and processes (Short 2023). A focus on public interest ensures greater emphasis on equity, but a challenge remains in regard to who determines public interest and what kinds of decisions need to be taken to ensure that determining public interest is not taken in arbitrary ways. In unpacking this concept, we introduce three elements: the existence of many publics, the prevalence of diverse interpretations, and the critical importance of deliberation and democratic debate.

First, the notion of public can be seen in and of itself as multifaceted and thinking about diverse publics within the EU is crucial. The concept is socially and culturally diverse and taking a pluralistic and polycentric approach recognises that there are overlapping interests, authorities, and communities that need to be accounted for, which will bring to the fore the need for cooperation and collaboration (Ostrom 2010). This is strongly advocated for all actors examined throughout the chapters of this document.

Second, public interest is subject to diverse interpretations. This study argues the need to strengthen the balance between: (1) those mechanisms that can be seen as top-down definitions of public interest data (e.g. HVDs represent documents held by a public sector body, the reuse of which is associated with important benefits for society, the environment and the economy[36]); and (2) the need to foster co-creation in the definition of public interest with the actual relevant public (see, e.g. Scharfbillig et al., 2021). For instance, regarding the latter case, thinking about the value of data through co-creation will require engaging with competing views where data can be seen to have economic, social, as well as embodied value (Russo Carrol et al. 2020). Taking these different viewpoints will entail that, in some cases, data is for extractive purposes to secure its economic value, but in other cases more data is not the answer and ceases to benefit people.

Third, concerning public interest data, deliberation and cooperation between different entities are required in terms of how data are created, what kinds of interests are represented, what kinds of use are undertake, and how data adapt and change according to the needs of the community. This entails that this definition itself should be subject to rigorous democratic processes to ensure that it is reflective of the needs of actors in a data ecosystem (Tarkowski et al 2022).

36. https://digital-strategy.ec.europa.eu/en/news/commission-defines-high-value-datasets-be-made-available-re-use

To do so, as mentioned above, **use cases, success stories and best practices can play a significant role**. These concepts go hand in hand, since they should inform public interest data. The challenge for the new legal framework is to put in place methodologies and processes to capture the constantly evolving notion of public interest and manage use cases effectively.

Use cases (both successful and not) are seen as a key driver for achieving a systemic vision of the GDDS, whereby a demand-driven perspective is effectively integrated. The concept of use-cases-driven data innovation is, however, not new. Still, challenges have emerged to-date in the successful implementation and long-term sustainability of such approaches, especially when use cases are promoted at their piloting stage. In addition, use cases are arguably difficult to monitor and manage, and simply capturing and showcasing these are not enough. This study proposes the following pathways:

1. **Building an observatory of use cases in the form of a living and agile resource** including both existing and new ones. This observatory can also evolve as a repository of best practices (extracted from the successful use cases themselves) to create data-enabled impact on the Green Deal;
2. **Emphasising the impact of such use cases as well as the data and the processes** that enabled their successful implementation, through the dissemination of impact stories and other initiatives, e.g. contests, awards etc.;
3. **Placing effort on upscaling successful use cases** through e.g. integrating these use cases into public procurement processes and schemes, or, more generally, to position these in a coherent scaling framework and roadmap. This could involve considering options to scale up, out, deep, and down as defined in social innovation literature.

It is important to ensure that the use cases remain reflective of the public's needs and continue to serve in their evolving interests. In other words, learnings from emerging and established uses cases should serve two purposes: (i) trigger sustainability and upscaling mechanisms; and (ii) taking the opportunity, given the outcomes achieved (and achievable), to continuously reflect on the notion of public interest and participate in the dynamic development of the concept.

Fourth and finally, the GDDS should also tackle the concept of private interest, which may or may not be aligned with the notion of public interest, consistent with the single data economy principle promoted across the EU. This study echoes the opinion that private interest data should not be tackled as an additional focus in parallel with public interest data. Rather, these should become part of the same system, emphasising the synergies that exist between them, with an emphasis on mechanisms to handle privacy, confidentiality and anonymity of data. Chapter 5 highlights several opportunities for integrating a business perspective into public interest data, or data for the common good. These are clustered around the following groups: profit incentives, Corporate Social Responsibility mechanisms, and reputation and trust gains (in this case by private actors and their offerings). There is, however, an important risk involved. This study reflects on how private interest data often leads to considering data as a corporate or strategic asset. This inevitably leads to businesses often operating based on self-interest, which may

not necessarily be aligned with the notion of public interest. Coupled with the fact that the private sector (when compared to the public sector) tends to possess higher capacities and resource availability, these scenarios often lead to power imbalances and differentials in the system. Chapter 5 also elaborates on B2G scenarios where private sector agencies and actors can produce data of public relevance (as already widely stressed across both the EU Strategy for Data and the Data Act).

8.2.2 The INSPIRE Directive would require a shift in focus from data availability to data use, participation, empowerment, and agency

Critical to the configuration of the INSPIRE Directive is an emphasis on being provider-centric, where a focus is on data discoverability, interoperability and accessibility, without actively integrating data use dynamics. Taking a Systemic Data Justice approach requires moving beyond availability to consider actors within a data ecosystem and the nature of interactions between them.

Such an approach investigates who is part of the data ecosystem (e.g. data producers, data users, data intermediaries) and what kinds of roles and responsibilities they each have. It requires examining what kinds of differential needs and capacities exist amongst actors, ranging from statistical capacity to human resource and financial capacity. From a data user standpoint, data literacy also emerges as an important consideration. Strengthening data ecosystems requires creating an enabling environment that would facilitate a culture of data sharing.

All chapters, in their own focus, highlight the current or potential existence of power imbalances in data ecosystems. This is primarily the case of those actors that are subject to the highest barriers for participation. The most evident example refers to citizens and communities whose perspectives are often absent or invisible, and as a result are often mere producers of data, rather than active shapers of its use and impact. As discussed in Chapter 6, the challenge is for these communities and individuals to be recognised as knowledge makers, i.e. providing experiential knowledge to complement existing data. Their participation and agency is too often limited by the societal (in)ability to recognise citizen-generated data as a valid form of knowledge creation or by the lack of resources and skills if compared to other actors. This discussion goes beyond citizens. For example, power imbalances can occur between businesses and governments, whereby the former leverage their typically greater capacities and resources, as explained in Chapter 5. These, if coupled with data being one of their strategic assets and the fact that businesses tend to act based on self-interest, leads to imbalances in negotiating power, also at the time of establishing highly advocated public-private partnerships.

Looking at the emergence of the GDDS, it becomes a priority to add clarity as to who is intended to participate in the Green Deal data ecosystem and how. Subsequently, capacity and resource gaps must be identified and addressed to the extent possible. This requires a new social contract that is premised on tackling layers of inequality from gender, race, class, ethnicity[37], as will strengthening civic engagement in shaping the GDDS. However,

37. https://www.un.org/sg/en/content/sg/statement/2020-07-18/secretary-generals-nelson-mandela-lecture-%E2%80%9Ctackling-the-inequality-pandemic-new-social-contract-for-new-era%E2%80%9D-delivered

this complex discussion goes beyond the scope of this report.

One of the most critical objectives of the GDDS should therefore be to foster **overcoming existing barriers for participation in this emerging ecosystem**, across technologies as well as both data provision and data use activities and processes. This translates into the need to foster agency of actors, especially those for whom existing data ecosystems result in being not inclusive or accessible altogether. As an example, the current open government data movements are often described as requiring substantial technical skills in order to participate. The result is typically the creation of an environment that is often exclusively designed for developers and tech-savvy individuals only. With respect to technologies, there is a need to ensure that these are within reach to all, and not only to a few. This means exploring and leveraging on certain reference protocols and their open source-based implementation. By doing so, technology is expected to become more accessible, also by actors that may not possess the capacity and capabilities of developing and maintaining these protocols by themselves. These include entities such as SMEs, professionals, NGOs, academics, and the public sector among others.

However, a reflection is required on what is meant by agency. First, the notion of agency should be extended from playing a role in the production of data, to including aspects of control over how data are used and shared. This reflection primarily considers the need to thoughtfully augment citizens' agency in the current EU data ecosystem. In this way, this study highlights the role that both Data Intermediaries Service Providers (DISPs) and Recognised Data Altruism Organisations (RDAOs) can play in this direction (see Chapters 3 and 4). If their remit is expanded, as argued in Chapter 3, DISPs have the potential to address current capacity gaps and power imbalances within and across the supply and demand of data. RDAOs can enhance individuals' and collectives' capacities, thus lowering an important barrier for participation, as well as reducing data and information asymmetries through, e.g. enabling individuals to (re)gain ownership of and decision-making power over the data they produce. As highlighted in chapter 4, the role of RDAOs is currently being shaped and the upcoming 'Rulebook' will add more clarity on the specific mechanisms these organisations can dispose of in order to address these challenges. At this moment, the early stage of RDAOs still leaves several questions open, such as the impact of compliance costs (e.g. GDPR) on their ability to enable agency for achieving the 'public interest' or common good.

As part of additional key enablers for agency and participation, **data literacy is seen as a critical building block** and this study argues for the need to foresee dedicated programmes and funding in this direction. Different stakeholders should be seen as the intended beneficiaries, such as public administrations, SMEs and other companies, and communities. This should not just align with data literacy principles but also include wider awareness initiatives about data culture. This means going beyond technical and specific governance mechanisms, through including literacy in new forms of data-driven innovation. Examples include: data altruism and more specific data management principles and practices, as well as how to exercise data stewardship. Depending on the stakeholder type, different capacity building

and literacy programmes should be promoted. For example, digital skills programmes may be prioritised for public administrations, and public-private partnerships have proven valuable in this direction. Consistent with the Systemic Data Justice perspective adopted here, it is critical that participation involves different publics, and in doing so, it does not become burdensome. Thereby, there should not only be clear guidance on how to participate, but also knowledge on what to expect. This requires integrating ideas of the CARE principles (see Chapter 7), where an emphasis is placed on empowerment that serves the collective benefit.

The new policy directions should address current power imbalances. The solution is seen primarily revolving around enabling new governance mechanisms and empowering those that are currently affected by the highest barriers to enter data ecosystems. This aspect is particularly relevant when considering the need of leaving no one behind for effectively achieving the Green Deal objectives. In this way, and connected to section 8.2.1 above, specific focus should be placed on empowering those ultimately affected by the Green Deal 'public interest', as they too often appear as the mere receivers of data-enabled products and services. Their participation should be enabled by a mix of incentives, roles, and guidelines. The important element is to involve appropriate actors across different sets of activities, such as:

- providing data subjects and data holders (individuals and entities) with means to monitor data use;
- putting in place instruments to enable feedback loops between end-users and developers and service providers;
- providing mechanisms for individuals and groups to register grievances in case of complaints regarding data production and use.

With respect to CGD, **mechanisms of accountability and legitimacy** should be in place to promote, under certain conditions, the right for communities to contribute to a given policy or issue, i.e. to produce valid knowledge (see Chapter 6). Examples of required conditions may revolve around the need to follow an acknowledged methodology as well as to put in place mechanisms to ensure quality and (scientific) validity of the contributions, i.e. to gain legitimacy to contribute to policymaking.

Several elements of this discussion relate to the concept of transparency, dominant in the open data landscape to-date (e.g. see Open Data Directive). Capacity building is also well embedded in this view.

8.2.3 INSPIRE should promote an environmental data value chain, moving towards an ecosystem approach

Moving from provider-centric to a data ecosystem-perspective necessitates strengthening capacities across the data value chain, to facilitate the participation of a diverse range of actors. This includes that technical considerations around data collection should be simplified and make data available to ensure broader participation beyond from public sector bodies. It also requires creating conditions to ensure that citizens and non-governmental actors are active in shaping the production, collection, and use of data, through data empowerment initiatives such as with data literacy, awareness, and ease of access to data.

Building an ecosystem around data relates to the notion of public interest data, where data are conceived of beyond a mere economic value, with importance is also given to the social value aspect. From a Systemic Data Justice standpoint, data are not just seen as a commodity, but rather incorporate different facets, including being about social relations and as a product of labour, and that data have a strongly embodied function (deSouza & Taylor forthcoming). In relation to green data, it is critical to acknowledge that data are also 'in the ground', and as indigenous sovereignty experts have argued data are 'living' (Russo Carrol et al 2021).

This understanding connects with the shift from a repository logic, as outlined previously in the INSPIRE Directive, to an **ecosystem logic, which would imply leveraging the strong community around INSPIRE** to become actively involved across the data value chain in shaping its future. In doing so, such a move would enable the identification of the purposes (plural) of the data ecosystem, and present it as an inclusive framework of different interests and capacities.

Across EU Member States, there is differential capacity in terms of statistical resources or human resources across official institutions as well as civil society. Consequently, participating in data ecosystems is a challenge, and is in fact a privilege until now. Efforts are therefore needed to address capacity gaps, by engaging with actors who are on the peripheries of data ecosystems to ensure engagement.

One example of this can emerge through standard setting, where standards for participation in data ecosystems present barriers. It is imperative that processes of standard setting are consultative and not top-down, ensuring that these **take into account Member States', citizens' and private businesses' concerns, needs, and capacities**. In addition to adopting FAIR principles to ensure that data is Findable, Accessible, Interoperable and Reusable, to account for power imbalances, it is important to adopt CARE principles, of Collective Benefit, Autonomy, Responsibility and Ethics to ensure that participants are not merely data providers, but active users, with sovereignty for their actions. Drawing from the CARE principles would entail a greater acknowledgment of how people wish to have their data governed, and what kinds of participation they would require to ensure that such governance is representative of their interest, and cognisant of their capabilities and values. Creating the conditions for such participation could result in the development of agile standards for instance on interoperability or usage rights to encourage data sharing.

It is also critical to consider concepts of data solidarity, which accounts for the horizontal as well as vertical relations that emerge through data. This approach recognises that hierarchies emerge in data collection, analysis, and use, and should be addressed to ensure that existing social inequalities are not exacerbated by data. **Incentives such as public security guarantees, financial compensation, or commercial guarantees**, as outlined in Chapter 2, play an important role in encouraging participation. They should be designed in a manner that does not reinforce the status quo in data ecosystems but instead creates a data culture that invites more diversity and plurality. In cases where incentives do not work, obligations involving furthering the public interest should also be considered.

8.2.4 INSPIRE should foster increased trust (both in the data and the processes) as a key intended outcome

The GDDS aims to be a trustworthy ecosystem for data sharing and use. When it comes to the intersection between environmental affairs and the GDDS, trust is understood as referring primarily to two aspects:

- Quality of data and datasets, which is related to the origins of the data. In this case, trust is generated on the validity of the data, the quality and thus usefulness. This is based on who generated and managed this data and, importantly, how. Trust in data ecosystems so far has primarily been thought of with respect to data, and improving data quality is seen as providing large parts of the solution.
- Certification processes enabled by existing legislation. This element may lead to trust in entities and the work they perform based on their recognition as a trusted entity. Certifications and accreditations can be mandatory or voluntary (also including those cases where recognition is built bottom up). Certificates are usually based on evidence and require human intervention and assessment of this evidence.

In summary, trust should be seen as the result of a combination of data and accountability of actions that actors perform on the data. As such, trust should be addressed and achieved not only in the data, but also and primarily in the processes and relationships embedded in the GDDS. However, when discussing trust building and maintenance, two considerations are needed. First, trust is an outcome that builds over time. It should therefore be thought of as a continuum between two extremes: full trust and distrust. While generating trust is the goal, situations of non-trust (or where distrust should be present) should also be contemplated. One risk of focusing on 'full trust only' could be that trust is achieved only in optimal, perfect situations and conditions, which are unlikely to occur. Second, trust is a multifaceted concept that may be subject to trade-offs among actors of the GDDS. Trust is not relevant in the same manner across all elements of the value chain or network. Existing directives and legislations, such as the DGA, enable trust through a form of institutionalisation of processes and a framework for recognition and certification. This report outlines **three antecedents of trust: (i) explainability, (ii) transparency, and (iii) accountability**. These three elements are particularly crucial in situations where the starting point is of distrust. These are therefore seen as desired outcomes to be addressed in a trust building process.

Data literacy is also advocated as an additional enabler of trust. **Awareness and the capacity of the general public to understand data and the surrounding processes** are critical and can foster trust, especially for those who distrust the current data system because of their inability to grasp and understand its principles, values, and functioning. Also, consistent with the key assumption that value from use is the main force driving the entire system, use cases are also seen as a critical enabler of trust. These further two elements (data literacy/awareness and use cases), are also relevant to gain more knowledge, experience, and confidence, leading to more active participation and critical thinking. All in all, these processes and building blocks can lead to both a situation where trust in a certain dataset or process is achieved or where these are meaningfully distrusted and

criticised. The discussion can be articulated across different actors and their related natures and scopes in the GDDS.

Achieving trust in the overall system is not limited to institutional trust, but also includes trust in private actors, data intermediaries, RDAOs, and communities. As highlighted above, the main mechanism currently available revolves around c**ertifications and accountability**. This includes accreditation bodies that are centralised and certify actors, processes, and data based on their compliance with data quality management systems, the use of standards and other more technical elements such as metadata, and security safeguards. In addition, audit bodies perform certifications that lead to trust (under the condition that the accreditation body is trusted itself). So, how can the GDDS enhance these? This study argues that the GDDS should leverage these existing environmental governance mechanisms (especially those which are mandatory) and digitalise them rather than defining new ones altogether. This digitalisation process is recommended as it will contribute to make the overall system more transparent, accountable and accessible, i.e. it will likely contribute to an increased trust in the GDDS.

In conclusion, the different chapters provide detailed insights and recommendations around trust, with respect to their specific focuses. For example, Chapter 2 highlights several incentives provided by existing legislations contributing to trust such as public security and commercial guarantees in data transfer agreements. Chapter 6 argues that, in the context of CGD, data quality and literacy lead to an increased legitimacy, which in turn enables an increased trust in the system whereby citizens can potentially become acknowledged knowledge makers. The chapters also highlight how legal uncertainty may present an issue such as in the case of the currently unclear governance and control mechanisms for RDAOs.

Finally trust can also be understood in terms of perceived sustainability in the future data provision. Chapter 4 argues how the nature of the challenges tackled within the Green Deal often requires a longitudinal approach to data collection and analysis. As a consequence, data users require a form of assurance that the data provided today will be sustainably provided in the future as well. While different governance mechanisms may exist depending on the type of data and the actors sharing it (e.g. Service Level Agreements in B2G scenarios), the GDDS should embed measures to augment the sense of perceived continuous data provision. This would ultimately contribute to augment trust in the data as a source of innovation, thus incentivising the investment of resources to reuse it in different ways and to actively participate in the GDDS.

8.2.5 INSPIRE should find an appropriate balance between centralised and decentralised governance mechanisms

It can be argued that data spaces in general ingrain both centralised and decentralised governance and data governance mechanisms and principles. While centralisation/ decentralisation can be defined in various ways within the context of data spaces, we adopt this terminology to refer to the extent to which the definition, oversight, and enforcement of data policies, standards, and rules are set by either a single (central) authority or distributed among multiple stakeholders. A parallel can be drawn with the concept of multilateral and non-multilateral governance schemes, which can be representative or of direct participation.

Given the scope of this report, the discussion refers to the role of the European Commission in this context.

By definition, the conversation around data spaces, their overarching vision, and (some of) the rules for participating in them are centrally designed, managed and controlled at the Commission level. As an example, the DGA defines the requirements and obligations for entities that want to act as data intermediaries (see Chapter 3) and RDAOs (see Chapter 4). However, data spaces have also been conceptualised under a decentralised approach. The Open DEI project defines four design principles and pillars for data spaces which include[38]: data sovereignty, a data level playing field, public/private governance, and decentralisation. This is interpreted with respect to the need to redistribute accountability, decision-making power and authority, away from a single entity, institution, or platform. In the context of data spaces, Page and Cecconi (2023) in their 'European data spaces and the role of data.europa.eu' report[39] argue that: *By promoting distribution [of power] and reducing reliance on a central point of control, decentralisation aims to enhance transparency, resilience and democratisation in various systems and applications* (p.6). As such, the co-existence of both centralised and decentralised principles and mechanisms is emerging in the conceptualisation of data spaces. The DGA provides some examples of this hybrid approach. It establishes data intermediaries as entities (or actors in the EU data economy) that: (1) prevent issues of centralisation through requiring structural separation between their contribution to the system and any other service they may provide; and (2) are governed by strict centralised requirements to ensure both avoiding conflicts of interest and neutrality (see Chapter 3). Further, the governance systems of RDAOs follow a similar federated approach; their registry is managed both centrally and by those Member States in which the specific organisation resides (see Chapter 4).

Overall, it is clear that both fully centralised and decentralised approaches are not suitable for the GDDS. Rather, a hybrid approach should be planned and adopted. The complex question to be addressed is which elements of the GDDS should be centralised and which ones should be decentralised.

The different chapters of this report provide insights into this debate. For example, Chapter 2, distinguishes between data governance, data space governance, and federated platforms. With respect to the latter, Chapter 3 reflects on data spaces compared with more common open (government) data platforms (i.e. usually fully centralised) and with data marketplaces. With respect to the federated approach of RDAOs described above, Chapter 4 describes a current problem of the authority delegated to national governments, which may lead to the creation of a scattered data altruism ecosystem, rather than an integrated network of RDAOs. In addition, situations of legal uncertainty may emerge in these cases such as in the interpretation of what may constitute 'common good' (and what not), which may vary from country to country.

To further contribute to this debate, this report acknowledges the benefits of both approaches and advocates for finding an appropriate, hybrid balance. **Centralisation efforts** should focus on interoperability among other issues. This means centralising

38. Open DEI received funding from the European Horizon 2020 programme for research, technological development and demonstration under grant agreement No 857065, https://www.opendei.eu.

39. Available at: https://data.europa.eu/sites/default/files/report/Data_Spaces_Panel_Report_EN.pdf

the effort of generating, establishing and maintaining catalogues of IT standards and the overarching IT reference architecture. Data discovery, search and other services should be centralised while access rules and storage should be decentralised closer to the source. Beyond interoperability, the benefits of centralised IT systems include increased and streamlined modifiability, operability, confidentiality, and integrity. From the opposite perspective, such an approach would substantially reduce the risk of generating a scattered ecosystem, i.e. the discrete sum of diverse national initiatives. Connected to this, there is the risk for RDAOs mentioned above. The expectation is that the upcoming rulebook and guidelines that will complement the existing text of the DGA will add clarity to these aspects in the currently federated approach taken.

Another aspect where this study supports centralisation of efforts relates to the overarching vision and strategic governance of the GDDS. The EC is, and is expected in the emerging GDDS, to take the lead in the definition of principles and values as well as strategic plans and overarching objectives of the Green Deal. Existing and forthcoming policies, directives and regulations provide the overarching goals and the 'rules of play' for data spaces, also including substantial funding mechanisms to address these. Importantly, these aspects foster alignment of accountability principles across the EU, which in turn is expected to increase trust in the system (see previous point).

On the other side of the spectrum, the benefits of **decentralised aspects and approaches** are typically related to an increased consistency with the diverse socio/legal/cultural and economic contexts that exist within the EU. For example, when discussing the agricultural data space, Falcão et al. (2023) argue that decentralised approaches are still desirable, because of legal (e.g. related to ensuring fair competition, cf. Kalmar et al., 2022) and organisational characteristics of the agricultural domain. Regarding the latter, agriculture as a sector is described as 'huge and scattered' and thus not suitable for a centralised approach.

The chapters in this study, also consistent with the notion of agency and empowerment, argue for more decentralised data governance mechanisms and for an infrastructure (both organisational and technological) that allows a wider distribution of power and decision rights regarding both data and processes at stake. Some aspects emerging from this study point towards decentralising access rights to data (specifically tackled in both the Data Act and the GDPR to address known challenges involving data ownership) as well as stewardship mechanisms. The latter include both the data as such and the initiatives built around datasets, e.g. those tackling the common good or public interest. As an additional example, Chapter 7 argues for decentralising access rights to data from Indigenous people and local communities.

Where the distinction between centralised and decentralised mechanisms remains more subtle is in the context of the need to facilitate, foster, or establish (depending on the level of centralisation) an overall multi-stakeholder and multi-level governance scheme. This report provides the perspectives of several actors in these conversations including businesses, citizens, communities, data intermediaries (within and beyond DISPs as defined in the DGA), RDAOs, and the public sector among others. The cross-chapter recommendations provided in terms of scope, agency and empowerment, and the reflections

on trust represent building blocks for fostering an effective, fair, just, and sustainable GDDS.

In conclusion, aspects of centralisation and decentralisation are multi-level and entail design choices. In general, we conclude that data governance rules, principles, and authentication services should assume a more centralised formulation, whereas data management should be decentralised 'at the source' (noting it refers to both the generation and use of data). Beyond general considerations, each specific use case will inform the most appropriate strategy to follow. This is similar to what happens in cloud-edge scenarios where each case may be suitable for one or the other depending on their need for computation and analytics. This last reflection points again towards the importance of capturing, monitoring, and disseminating use cases. This would allow efforts to provide a central repository of success (and failure) cases together with their recommended (best) practices as a dynamic, living resource. Users can then be actors ranging from enterprise to self-governing communities that act in a decentralised manner, driven by central rules and principles, while promoting and building one single EU market for data.

8.3 Summary and conclusions

This final chapter highlights key policy recommendations for the revision of the INSPIRE Directive, in the context of the common European Green Deal data space (GDDS) and towards a more sustainable and fair data ecosystem. It introduces the concept of **'Systemic Data Justice'** which entails a focus on equity, accountability, and fair representation to achieve stronger links between the supply and demand of data for a more effective and sustainable data economy.

The policy recommendations are grouped in **five main cross-cutting aspects**. The first point emphasises the need for a fundamental shift in the purpose of INSPIRE from public sector data to public interest data. This shift entails integrating a demand-driven perspective and exploring synergies between public and private interest data. The second recommendation underscores the importance of moving from a focus on data availability to data use, participation, and agency. It addresses power imbalances and advocates for a more inclusive and accessible data ecosystem, emphasising the need to empower actors across the data value chain, including data producers, users, and intermediaries.

Third, we highlight the necessity to promote an environmental data value chain, moving towards a data ecosystem centred approach. This approach aims to foster a data sharing and use economy that goes beyond a repository logic to ensure that the EU is a competitive digital market while taking into account societal inequalities and differentiated impacts in the green transition. The fourth recommendation emphasises the need to foster increased trust as a key intended outcome. This involves focusing on data quality, certification processes, and data literacy to ensure that the data shared within the GDDS are trustworthy and reliable. Lastly, the final point advocates for finding an appropriate balance between centralised and decentralised governance mechanisms. It suggests combining central governance rules with a decentralised approach to data management, informed by use cases.

Table (9) provides an overview of these key recommendations, their main implications and actionable points. In addition, it outlines the possible entities that could be responsible for implementing these actions, including various actors from the ecosystem, such

as data collectors and providers, data intermediaries and altruism organisations, Member States, and key entities from the European Commission. In particular, this proposal considers a potential scenario with a Commission Decision for the establishment of an expert advisory group with a focus on public sector Green Deal data. We refer to this new body as the **'Green Data Advisory Board'**, to be formed by representatives from national authorities, academia, industry, civil society, and other stakeholders such as standardisation bodies (OGC, ISO), leveraging and building upon the existing INSPIRE community.

Finally, it is important to highlight that the recommendations captured in this document are also **highly relevant for other data spaces**. On the one hand, the horizontal and cross-cutting nature of Green Deal data entails the need for consistent data governance approaches across data spaces that ensure no person or place is left behind. On the other hand, the relevance of adopting a Systemic Data Justice approach goes beyond Green Deal data only; it is instrumental to ensure that any data ecosystem is both inclusive and fair, addressing power imbalances and enabling more accessible and transparent data sharing and use practices. This will ensure not only wider participation in the data economy such that the data use can facilitate the empowerment of people and communities.

TABLE 9.

Policy recommendations for the revision of the INSPIRE Directive in light of the common European Green Deal data space.

Recommendation	Implications	Actionable points	Responsible / involved entities
i. In the context of GDDS, INSPIRE would require a shift in purpose from public sector data to public interest data.	• Greater focus on data use (integrate a demand-driven perspective). • Exploring and exploiting synergies between public and private interest data. • Prioritise 'mission first' actions over 'data first' ones.	• **Encourage greater openness and participation from non-traditional actors** by including novel data sources and removing legal and technical barriers through the development of agile frameworks that can encourage other actors such as academia, businesses or citizen groups to participate in data production, collection, sharing, and use. To do so, legitimacy and recognition at the institutional level would be needed. • **The GDDS should promote and maintain an observatory of use cases and an associated repository of best practices** as a living resource, which promotes both the exploration of new solutions and the scaling-up and replication of existing applications and practices (e.g. through procurement). • **Co-create the definition of public interest data**, considering their economic, social, and environmental value and potential uses.	• **Green Data Advisory Board and EC Directorates General** working with environmentally-relevant common European data spaces (i.e. DG ENV, CNECT, DIGIT, CLIMA, RTD, AGRI, JRC) • Initiative to be developed with support from **REA Funding (e.g. Horizon Europe)** and in-kind contributions from Member States. • **Green Data Advisory Board facilitated by the DSSC**

Recommendation	Implications	Actionable points	Responsible / involved entities
ii. INSPIRE would require a shift in focus from data availability to data use, participation, and agency.	• Responding to differential needs and capacities amongst the actors in the GDDS ecosystem (data producers, data users, data intermediaries).	• **Identify and address capacity and resource gaps** among those who are intended to participate in the Green Deal data ecosystem. This comprises providing dedicated programmes and funding to improve digital literacy and awareness of various actors across the supply and demand sides of data – from public administrations to SMEs and communities. • **Provide a mix of incentives to enhance participation**, such as: (i) means to monitor data use from data subjects and data holders, (ii) instruments to enable feedback loops between end-users and developers/service providers, (iii) mechanisms to register grievances regarding data production and use, (iv) means to ensure accountability and legitimacy of citizen-generated data, and (v) promoting rewards for best practices (e.g. labelling, annual rewards, etc.). • **Provide a clear and transparent accountability framework** for entities/individuals who are formally recognised within the GDDS with respect to the role they perform and the associated decision rights, as well as at the community level	• **MFF funding allocation linked to bottom-up** initiatives from Member States (local, regional, and national public administrations) and non-public associations. • **Data intermediaries and data altruism organisations** • Green Data Advisory Board (guidance role) • **Green Data Advisory Board**
iii. INSPIRE should promote an environmental data value chain, moving towards an ecosystem approach.	• Moving from provider-centric to data ecosystem-centric.	• **Simplify technical considerations around data collection** across the data lifecycle to broaden participation beyond public sector bodies, while ensuring interoperability is main-tained. • **Set standards based on FAIR[40] principles and in a consultative manner**, considering needs, concerns, and capacities of Member States, local authorities, citizens and private businesses. • **Embed CARE[41] principles, data solidarity, and sovereignty** aspects to ensure participation is fair, inclusive, and effective.	• **Green Data Advisory Board and online communities** (e.g. through GitHub), building upon lessons learned and best practices from the INSPIRE Community. • **Data collectors and providers**

40. Findable, Accessible, Interoperable, and Reusable.

41. Collective benefit, Autonomy, Responsibility and Ethics.

Recommendation	Implications	Actionable points	Responsible / involved entities
iv. INSPIRE should foster increased trust (both in the data and the process) as a key intended outcome.	• Providing a trustworthy system for data sharing and use comprising institutions, private actors, data intermediaries, altruism organisations, and communities.	• **Enhance and digitalise certifications and accreditations** to increase trust in entities processing data. It can be mandatory or voluntary, and include recognition built in a bottom-up manner. • **Enrich datasets with quality assurance mechanisms and openly disclose information** on how the specific data has been produced, when, and by whom. • **Digitalise existing environmental governance mechanisms** increasing transparency and accessibility. • **Increase data literacy and awareness of data impacts** to gain more knowledge, experience, and confidence in the data and their processes.	• **SIMPL middleware platform** • **Data intermediaries and data altruism organisations** • **Data collectors and providers** • **Observatory of uses cases** (mentioned above)
v. INSPIRE should find an appropriate balance between centralised and decentralised governance mechanisms.	• Leveraging the benefits of centralised IT systems, including increased and streamlined modifiability, operability, confidentiality, and integrity. • Ensuring adaptation and compliance with a variety of decentralised legal, social, and cultural European contexts.	• **Embed a federated approach, where central actions (from the European Commission) can focus on interoperability**, (including generating, establishing, and maintaining IT standards and the overarching IT reference architecture), data discovery, search, and other services, while access rules and storage are decentralised, closer to the source. • **Define a central, unifying vision** including strategic objectives, values, and governance principles. • **Decentralise access rights, ownership, and stewardship** of both data and data-driven initiatives. • Facilitate, foster, or establish (depending on the level of centralisation) an overall **multi-stakeholder and multi-level governance scheme.** • **Combine central governance rules, principles, and authentication services** for data, **with a decentralised approach to data management** 'at the source' (both generation and use) informed by use cases.	• **Green Data Advisory Board** • **EC Directorates General** working with environmentally-relevant common European data spaces (i.e. DG ENV, CNECT, DIGIT, CLIMA, RTD, AGRI, JRC) • **Data intermediaries and data altruism organisations** (new business models leveraged by the Data Act) • **Green Data Advisory Board** • **Orchestrated by the DSSC**, and supported by **the Green Data Advisory Board**

Source: own elaboration.

REFERENCES

Ada Lovelace Institute., *Participatory data stewardship*, 2021. Retrieved from https://www.adalovelaceinstitute.org/report/participatory-data-stewardship/

Alemanno, A., *Big Data for Good: Unlocking Privately-Held Data to the Benefit of the Many*. European Journal of Risk Regulation, 9(2), 2018, pp. 183–191. doi:10.1017/err.2018.34

Ampumuza, C., *Living with Gorillas? Lessons from Batwa-Gorillas' Convivial Relations at Bwindi Forest, Uganda*, Conservation and Society, 20(2), Sage Publications, 2022, p. 69. https://doi.org/10.4103/cs.cs_30_21

Baack, S., 'Civic Tech at mySociety: How the Imagined Affordances of Data Shape Data Activism', *Krisis: Journal for Contemporary Philosophy*, No. 1, 2018.

Ballantyne, A., 'How should we think about clinical data ownership?', *Journal of Medical Ethics*, Vol. 46, 2020, pp. 289–294.

Balázs, B., P. Mooney, E. Nováková, L. Bastin, and J. Jokar Arsanjani, 'Data Quality in Citizen Science', in K. Vohland, A. Land-Zandstra, L. Ceccaroni, R. Lemmens, J. Perelló, M. Ponti, R. Samson, and K. Wagenknecht (eds.), *The Science of Citizen Science*, Springer, Cham, 2021, pp. 139–157.

Baloup, J. Bayamliyoglu, E., Benmayor, A., Ducuing, C., Dutkiewicz, L., Lalova, T., Miadzvetskaya, Y and Peeters, B. *White Paper on the Data Governance Act*, CiTiP Working Paper, KU Leuven Centre for IT & IP Law – imec, 2021.

Battiste, M., *Research Ethics for Protecting Indigenous Knowledge and Heritage: Institutional and Researcher Responsibilities*, in N. Denzin, Y. Lincoln, & L. T. Smith (Eds.), Handbook of Critical and Indigenous Methodologies, SAGE Publications, Los Angeles, 2008, pp. 497–510.

Berners-Lee, T. *Linked Data - Design Issues*, 2006. http://www.w3.org/DesignIssues/LinkedData.htm

Berti Suman, A., 'Challenging Risk Governance Patterns through Citizen Sensing: The Schiphol Airport Case', *International Review of Law, Computers & Technology*, Vol. 32, No. 1, 2018, pp. 155–173.

Berti Suman, A., *Citizen Sensing from a Legal Standpoint: Legitimizing the Practice under the Aarhus Framework*, Journal for European Environmental & Planning Law, 18(1-2), Brill, 2021, pp. 8–38. https://doi.org/10.1163/18760104-18010003

Berti, A. *Civic monitoring as a proactive form of public engagement*, EU Community of Practice of the Competence Centre on Participatory and Deliberative Democracy, Accessed on 18th June 2024 https://cop-demos.jrc.ec.europa.eu/blog/civic-monitoring-proactive-form-public-engagement

Berti Suman, A., Balestrini, M., Haklay, M., & Schade, S., *When Concerned People Produce Environmental Information: A Need to Re-Think Existing Legal Frameworks and Governance Models?*, Citizen Science: Theory and Practice, 8(1), Ubiquity Press, 2023, p. 10. https://doi.org/10.5334/cstp.496

Berti Suman, A., Schade, S., and Abe, Y. 'Exploring legitimization strategies for

contested uses of citizen-generated data for policy'., in B.J. Richardson (ed.), *From Student Strikes to the Extinction Rebellion*. Edward Elgar, Cheltenham, UK, 2020, pp. 74-102.

Borgogno, O. & Colangelo, G., *Data sharing and interoperability: Fostering innovation and competition through APIs*, Computer Law & Security Review, 35(5), Elsevier, 2019, pp. 1-17. https://doi.org/10.1016/j.clsr.2019.03.008.

Brevini, B., *Is AI Good for the Planet?*, Polity, 2021.

Brunton, F., and H. Nissenbaum, *Obfuscation: A User's Guide for Privacy and Protest*, MIT Press, Cambridge, 2015.

Bonney, R., *Expanding the Impact of Citizen Science*, BioScience, 71(5), Oxford University Press, 2021, pp. 448–451. https://doi.org/10.1093/biosci/biab041

Burgess, H.K., L.B. DeBey, H.E. Froehlich, N. Schmidt, E.J. Theobald, A.K. Ettinger, J. HilleRisLambers, J. Tewksbury, and J.K. Parrish, 'The Science of Citizen Science: Exploring Barriers to Use as a Primary Research Tool', *The Role of Citizen Science in Biological Conservation*, Vol. 208, April 1, 2017, pp. 113–120.

Carballa Smichowski, B., *Alternative Data Governance Models: Moving Beyond One-Size-Fits-All Solutions*, Intereconomics, 54(4), Springer, 2019, pp. 222–227. https://doi.org/10.1007/s10272-019-0828-x

Carlsson B. and Stankiewicz R., *On the nature, function and composition of technological systems*, Journal of Evolutionary Economics, 1991, 1(2), 93–118.

Carovano, G. and Finck, M., 'Regulating data intermediaries: the impact of the Data Governance Act on the EU's data economy', *Computer Law & Security Review*, Vol. 50, 105830, 2023 https://doi.org/10.1016/j.clsr.2023.105830

Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Hudson, M., *The CARE Principles for Indigenous Data Governance*, Data Science Journal, 19(1), Ubiquity Press, 2020. https://doi.org/10.5334/dsj-2020-043

Carroll, S. R., Herczog, E., Hudson, M., Russell, K., & Stall, S., *Operationalizing the CARE and FAIR Principles for Indigenous data futures*, Scientific Data, 8(1), Springer Nature, 2021. https://doi.org/10.1038/s41597-021-00892-0

Carroll, S. R., Kukutai, T., Walter, M., Figueroa-Rodríguez, O., Walker, J., & Axelsson, P., Indigenous data sovereignty, in T. Davies, S. B. Walker, M. Rubinstein, & F. Perini (Eds.), *The state of open data: histories and horizons*, African Minds and International Development Research Centre, Cape Town and Ottawa, 2019, pp. 300–319.

Carroll, S. R., Rodriguez-Lonebear, D., & Martinez, A., *Indigenous Data Governance: Strategies from United States Native Nations*, Data Science Journal, 18(1), Ubiquity Press, 2019. https://doi.org/10.5334/dsj-2019-031

Cetl, V., F. Penninga, J. Robbrecht, M. Lutz, A. Kotsev, and M. Minghini, *INSPIRE, a Public Sector Contribution to the European Green Deal Data Space: A Vision for the Technological Evolution of Europe's Spatial Data Infrastructures for 2030*, Publications Office of the European Union, Luxembourg, 2021.

Chae, B.K., *A General framework for studying the evolution of the digital innovation ecosystem: The case of big data*. International Journal of Information Management, 2019, 45, pp. 83-94.

Charvat, K., Bye, B.L., Mildorf, T., Berre, A.J. & Jedlicka, K., *Open data, VGI and citizen observatories INSPIRE hackathon*, International Journal of Spatial Data Infrastructures Research, 13, 2018, pp. 109-130.

Chignard, S. & Glatron, M., *Data collaborations at a local scale: Lessons learnt in Rennes (2010–2021)*, Data & Policy, 5, Cambridge University Press, 2023, e20. doi:10.1017/dap.2023.16

Comandé, G. & Schneider, G., *It's time: Leveraging the GDPR to shift the balance towards research-friendly EU data spaces*, Common Market Law Review, 59 (3), Kluwer Law International, 2022, pp. 739-776.

Corburn, J., *Street Science: Community Knowledge and Environmental Health Justice*, MIT Press, Cambridge, MA, 2005.

Corry, O. and Reiner, D. *Protests and policies: How radical social movement activists engage with climate policy dilemmas*, Sociology, Vol.55:1, 2021, pp. 197–217.

Costanza-Chock, S., *Design justice. Community-led practices to build the worlds we need*. Cambridge, MA: MIT Press, 2020.

Couldry, N., and U.A. Mejias, *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism, Culture and Economic Life*, Stanford University Press, Stanford, California, 2019.

Couldry, N., and A. Powell, 'Big Data from the Bottom Up', Big Data & Society, Vol. 1, No. 2, 2014, pp. 1–5.

Cox, L., and C. Flesher Fominaya, 'Movement Knowledge: What Do We Know, How Do We Create Knowledge and What Do We Do with It?', *Interface: A Journal For and About Social Movements*, Vol. 1, No. 1, 2009, pp. 1–20.

Cukier, K., and V. Mayer-Schoenberger, 'The Rise of Big Data: How It's Changing the Way We Think about the World', *Foreign Affairs*, Vol. 92, No. 3, 2013, pp. 28–40.

Currie, M., and B.S. Paris, 'Back-Ups for the Future: Archival Practices for Data Activism', *Archives and Manuscripts*, Vol. 46, No. 2, 2018, pp. 124–142.

Curry, E., *Real-time linked dataspaces: Enabling data ecosystems for intelligent systems*, Springer Nature, 2020, p. 325.

Daly, A., Private Power, *Online Information Flows and EU Law: Mind the Gap*, Hart Publishing, 2016.

Daly, A., Devitt, S. K., & Mann, M., *Good data*, Institute Of Network Cultures, Amsterdam, 2019.

Data Portal., *Analytical Report 12: Business-to-Government Data Sharing*, 2020. https://data.europa.eu/sites/default/files/analytical_report_12_business_government_data_sharing.pdf

Davis, J.L., and J.B. Chouinard, 'Theorizing Affordances: From Request to Refuse', *Bulletin of Science, Technology & Society*, Vol. 36, No. 4, 2016, pp. 241–248.

de Souza, S. P. & Taylor, L., (2024, September 3). Rebooting the global consensus: norm entrepreneurship, data governance and the majority world. https://doi.org/10.31235/osf.io/a2utr

Della Porta, D., and E. Pavan, 'Repertoires of Knowledge Practices: Social Movements in Times of Crisis', *Qualitative Research in Organizations and Management*, Vol. 12, No. 4, 2017, pp. 297–314.

Dencik, L., Jansen, F. & Metcalfe, *A conceptual framework for approaching social justice in an age of datafication*, DATAJUSTICE project, 2018. https://datajusticeproject.net/2018/08/30/a-conceptual-framework-for-approaching-social-justice-in-an-age-of-datafication/ (Accessed: 27 June 2024).

Dencik, L. and Sanchez-Monedero, J. *Data justice*, Internet Policy Review, 11(1), 2022. Available at: https://policyreview.info/articles/

analysis/data-justice (Accessed: 11 March 2023).

Deshmukh, R.A., Collarana, D., Gelhaar, J., Theissen-Lipp, J., Lange, C., Arnold, B.T., Curry, E. and Decker, S., *Challenges and Opportunities for Enabling the Next Generation of Cross-Domain Dataspaces*. In The Second International Workshop on Semantics in Dataspaces, co-located with the Extended Semantic Web Conference, May 2024.

Dietrich, M. and Gutiérrez David, M. (2024), *Green Deal Data Space and Foundation and its Community of Practice (GREAT) D4.2 Final Governance Requirements and Endorsed Governance Scheme*, https://www.greatproject.eu/wp-content/uploads/2024/04/D4.2-Final-Governance-Requirements-and-Endorsed-Governance-Scheme.pdf

Dinh, T., Husmann, A. & Melloni, G., *Corporate Sustainability Reporting in Europe: A Scoping Review*, Accounting in Europe, 20(1), Routledge, 2023, pp. 1–29. https://doi.org/10.1080/17449480.2022.2149345

Ducuing C. Data governance for the ecological transition: An infrastructure perspective. *Review of European Community and International Environmental Law* (2024), 1-12. https://doi.org/10.1111/reel.12547

'D-NOSES', *Science for Change*, 2017. https://scienceforchange.eu/en/project/d-noses/.

Hossain, N.U.I., Dayarathna, V.L., Nagahi, M. and Jaradat, R., *Systems thinking: A review and bibliometric analysis*, Systems, 2020, 8(3), p. 23.

Eckardt, M., Kerber, W. Property rights theory, bundles of rights on IoT data, and the EU Data Act. Eur J Law Econ, 2024. https://doi.org/10.1007/s10657-023-09791-8

EGD Prioritised Data Sets and Gaps (Initial Inventory plus Phase 1 Reference Use Cases), GREAT Consortium, 2024.

Espinosa Apraez, B., *Reconsidering the Public-Private Data Dichotomy in the European Union's Data Sharing Policies*, European Journal of Law and Technology, 12 (1), 2021. https://ejlt.org/index.php/ejlt/article/view/825

European Commission, Directorate-General for Environment., *Citizen science for environmental policy: development of an EU-wide inventory and analysis of selected practices*, Publications Office of the European Union, 2018. Retrieved from https://data.europa.eu/doi/10.2779/961304

European Commission, *Towards a common European data space*, 2018. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=COM:2018:0232:FIN

European Commission, *The European Green* Deal, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions (COM(2019)640 final, 2019.

European Commission, *A European strategy for data*, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions COM(2020) 66 final, 2020. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0066%20&from=EN.

European Commission, *The European Green Deal: A growth strategy that protects the climate*, 2020a. Accessed on June 27, 2024 at: https://ec.europa.eu/stories/european-green-deal

European Commission *Staff Working Document. Best Practices in Citizen Science for Environmental Monitoring*, Pub. L. No. 9973/20 (2020), 2020b .

European Commission, *Information session on a preparatory action for the common European Green Deal Data Space under the Digital Europe Programme' (DIGITAL)*, December 15th, 2021. Retrieved from: https://digital-strategy.ec.europa.eu/en/events/information-session-preparatory-action-common-european-green-deal-data-space-under-digital-europe.

European Commission, *Commission staff working document evaluation of Directive 2007/2/EC establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)*, 2022. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022SC0195.

European Commission, Digital Europe Programme (DIGITAL), https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/digital/wp-call/2022/call-fiche_digital-2022-cloud-ai-03_en.pdf, 2022a. Accessed on January 18, 2024.

European Commission, *Commission Staff Working Document on Common European Data Spaces*, SWD(2022) 45 final, 2022b.

European Commission, *Study on the Open Data Directive, Data Governance and Data Act and Their Possible Impact on Research*, Directorate General for Research and Innovation, Publications Office, LU, 2022c.

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, 2023, L 19: 43–75. Available at: https://eur-lex.europa.eu/eli/reg_impl/2023/138/oj.

European Commission, *Call for Evidence for an Impact Assessment of the INSPIRE Directive (Directive 2007/02/EC)*, 2024. Available at: https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13170-GreenData4All-updated-rules-on-geospatial-environmental-data-and-access-to-environmental-information_en.

European Data Portal, *Analytical Report n 12: Business-to-Government Data Sharing*, 2020. https://data.europa.eu/sites/default/files/analytical_report_12_business_government_data_sharing.pdf

European Green Deal Coalition, *Net Carbon Impact Assessment Methodology for ICT Solutions*, 2024. https://www.greendigitalcoalition.eu/assets/uploads/2024/04/EGDC-Net-Carbon-Impact-Assessment-Methodology-for-ICT-Solutions.pdf

European Union, *Directive 2003/4/EC of the European Parliament and of the Council of 28 January 2003 on public access to environmental information and repealing Council Directive 90/313/EEC*, Official Journal of the European Union, 2003, L 041: 26–32. (Freedom of Access to Information Directive). Available at: https://eur-lex.europa.eu/eli/dir/2003/4/oj

European Union, *Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2017 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)*, Official Journal of the European Union, 2007, L 108: 1–14. Available at: https://eur-lex.europa.eu/eli/dir/2007/2/2019-06-26.

European Union, *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)*, Official Journal of the European Union, 2016, L 119:

1–88. Available at: https://eur-lex.europa.eu/eli/reg/2016/679/oj.

European Union, *Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information*, Official Journal of the European Union, 2019, L 172: 56–83. Available at: https://eur-lex.europa.eu/eli/dir/2019/1024/oj.

European Union, *Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act)*. Available at: https://eur-lex.europa.eu/eli/reg/2022/868/oj

European Union, *Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act)*. Available at: https://eur-lex.europa.eu/eli/reg/2023/2854

Falcão, R., Matar, R., Rauch, B., Elberzhager, F. & Koch, M., *A Reference Architecture for Enabling Interoperability and Data Sovereignty in the Agricultural Data Space*, Information, 14(3), MDPI, 2023, p. 197.

Fink, A. *Data cooperative*. Internet Policy Review, Vol. 13(2), 2024.

Finck, M. and Mueller, M., *Access to data for environmental purposes: setting the scene and evaluating recent changes in EU data* law, *Journal of Environmental Law*, Vol. 35, 2023, pp. 109–131. https://doi.org/10.1093/jel/eqad006

Fishkin, J., *When the people speak: Deliberative democracy and public consultation*. Oxford: Oxford University Press, 2009.

Fricker, M., *Epistemic Injustice: Power and the Ethics of Knowing*., Oxford University Press, Oxford, 2007.

Frank Biermann, *'Earth system governance' as a crosscutting theme of global change research*, Global Environmental Change, Volume 17, Issues 3–4, 2007, Pages 326-337.

Fritz, S., L. See, T. Carlson, M. (Muki) Haklay, J.L. Oliver, D. Fraisl, R. Mondardini, et al., 'Citizen Science and the United Nations Sustainable Development Goals', *Nature Sustainability*, Vol. 2, No. 10, October 1, 2019, pp. 922–930.

Fritzenkötter, J., Hohoff, L., Pierri, P., Verhulst, S.G., Young, A., and Zacharzewski, A. (2022), *Governing the environment related data space*. SSRN Electronic Journal, 2022 http://dx.doi.org/10.2139/ssrn.4250166.

Gabrys, J., 'Citizen Sensing, Air Pollution and Fracking: From 'Caring about Your Air' to Speculative Practices of Evidencing Harm', *The Sociological Review Monographs*, Vol. 65, No. 2, 2017, pp. 172–192.

Gabrys, J., H. Pritchard, and B. Barratt, 'Just Good Enough Data: Figuring Data Citizenships through Air Pollution Sensing and Data Stories', *Big Data & Societ*y, Vol. 1, No. 14, 2016.

Galbiati, R. and Vertova, P. (2014), 'How laws affect behavior: obligations, incentives and cooperative behavior', *International Review of Law and Economics*, 38, 48–57. https://doi.org/10.1016/j.irle.2014.03.001

Geertz, C., *The Interpretation of Cultures: Selected Essays*, Basic Books, New York, 1973.

Getschmann, C., and F. Echtler, 'DesPat: Smartphone-Based Object Detection for Citizen Science and Urban Surveys', *I-Com*, Vol. 20, No. 2, 2021.

Gieryn, T.F., 'Boundary-Work and the Demarcation of Science from Non-Science: Strains and Interests in Professional Ideologies of Scientists', *American Sociological Review*, Vol. 48, No. 6, 1983, pp. 781–795.

Granell, C., P. Mooney, M. Rieke, F. Ostermann, J. van den Broecke, A. Sarretta, S. Verhulst, et al., *Emerging Approaches for Data-Driven Innovation in Europe: Sandbox Experiments on the Governance of Data and Technology*, Publications Office of the European Union, Luxembourg, 2022.

GREAT Project. (n.d.). *Glossary*. GREAT Project. Retrieved on July 31, 2024, from https://www.greatproject.eu/glossary/

Gurstein, M. B., 'Open data: Empowering the empowered or effective data use for everyone?', *First Monday*, Vol. 16(2), 2011.

Gutierrez, M., *Data Activism and Social Change*, Palgrave MacMillan, Cham, 2018.

Haklay, M., A. König, F. Moustard, and N. Aspee, 'Citizen Science and Post-Normal Science's Extended Peer Community: Identifying Overlaps by Mapping Typologies', *Futures*, Vol. 150, June 1, 2023, p. 103178.

Haklay, M., 'Citizen Science and Volunteered Geographic Information: Overview and Typology of Participation', in D. Sui, S. Elwood, and M. Goodchild (eds.), *Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice*, Springer Netherlands, Dordrecht, 2013, pp. 105–122.

Hintz, A., L. Dencik, and K. Wahl-Jorgensen, *Digital Citizen*ship in a Datafied Society, Polity, Cambridge, 2018.

Hall, M.A. and Wright, R. F., *'Systematic content analysis of judicial opinions'*, Wake Forest Univ. Legal Studies Paper No. 913336, California Law Review, Vol. 96, 2008, 1st Annual Conference on Empirical Legal Studies Paper, (June 30, 2006). Available at SSRN: https://ssrn.com/abstract=913336

Hansen, J; Wilson, P. Verhoeven, E. Kroneman, M.; Kirwan, M. Verheij, M. van Veen, E., *Assessment of the EU Member States' rules on health data in the light of GDPR*, Luxembourg: Publications Office of the European Union, 2021.

Heeks, R. & Renken, J., *Data justice for development: What would it mean?*, Information Development, 34(1), SAGE Publications, 2018, pp. 90–102. https://doi.org/10.1177/0266666916678282

High-Level Expert Group on Business-to-Government Data Sharing., *Towards a European strategy on business-to-government data sharing for the public interest*, 2020. https://data.europa.eu/doi/10.2759/731415

Ho, C. & CHUANG, T., *Governance of Communal Data Sharing*, in A. Daly, S. K. Devitt, & M. Mann (Eds.), Good Data, Institute of Network Cultures, Amsterdam, 2019, pp. 202–213.

Holton, J., *The Private Sector to the Public Sector's Rescue: Why Private Organisations Enter Crisis Response Data Collaboratives*, University of Leiden, 2018. https://studenttheses.universiteitleiden.nl/handle/1887/58629

Hsieh, H.-F., and Shannon, S. E. (2005), 'Three approaches to qualitative content analysis', *Qualitative Health Research*, 15(9), 1277–1288. https://doi.org/10.1177/1049732305276687

Ibarra, I.A., Goff, L., Jimenez Hernandez, D. Lanier, J. and Weyl, E. G., *Should We Treat Data as Labor? Moving Beyond 'Free'*, Conference Proceeding, 2018.

IPCC., *Climate Change 2022: Impacts, Adaptation and Vulnerability. Working Group II Contribution to the IPCC Sixth Assessment Report*, Cambridge University Press, Cambridge, UK and New York, NY, USA, 2022.

Jasanoff, Sheila, ed., 'States of Knowledge: The Co-Production of Science and the Social Order', Routledge, New York, 2004.

Kalkar, U. and González Alarcón, N., Facilitating Data Flows through Data Collaboratives A Practical Guide to Designing Valuable, Accessible, and Responsible Data Collaboratives, The GovLab and the Inter-American Development Bank, 2023.

Kalmar, R.; Rauch, B.; Dörr, J.; Liggesmeyer, P., *Agricultural Data Space*, Des. Data Spaces, 2022, pp. 279–290.

Kerber, W., 'Governance of IoT data: why the EU Data Act will not fulfill Its objectives', *GRUR International*, 72(2), pp. 120–135, 2023 https://doi.org/10.1093/grurint/ikac107

Kerber, W., *EU Data Act: Will new user access and sharing rights on IoT data help competition and innovation?*, Journal of Antitrust Enforcement, Oxford University Press, 2024, jnae011. https://doi.org/10.1093/jaenfo/jnae011

Kitchin, R., 'Big Data, New Epistemologies and Paradigm Shifts', Big Data & Society, 2014.Lähteenoja, V., *What are 'personal data spaces'?*. In Companion Proceedings of the ACM Web Conference 2023 (WWW '23 Companion), April 30-May 04, 2023, Austin, TX, USA. ACM, New York, NY, https://doi.org/10.1145/3543873.3587656

Kitsios, F. & Kamariotou, M., *Open data hackathons: an innovative strategy to enhance entrepreneurial intention*, International Journal of Innovation Science, 10(4), 2018, pp. 519-538.

Klein, T. & Verhulst, S., *Access to new data sources for statistics: Business models and incentives for the corporate sector*, OECD Statistics Working Papers, No. 2017/06, OECD Publishing, 2017. https://doi.org/10.1787/9a1fa77f-en

Kotsev, A., J. Escriu, and M. Minghini, *Beyond INSPIRE. Perspectives on the Legal Foundation of the European Green Deal Data Space*, European Commission, 2023.

Kotsev A., Minghini M., Cetl V., Penninga F., Robbrecht J., Lutz M., *INSPIRE – A Public Sector Contribution to the European Green Deal Data Space. A vision for the technological evolution of Europe's Spatial Data Infrastructures for 2030*, EUR 30832 EN, Publications Office of the European Union, Luxembourg, 2021, ISBN 978-92-76-41564-0, doi:10.2760/8563, JRC126319.

Kukutai, T., Carroll, S. R., & Walter, M., *Indigenous data sovereignty*, in D. Mamo (Ed.), The Indigenous World 2020, IWGIA, Copenhagen, 2020, pp. 654–662.

Lalova-Spinks, T.; Meszaros, J. and Huys, I., 'The application of data altruism in clinical research through empirical and legal analysis lenses'. In *Front. Med.*, Vol. 10:1141685, 2023.

Lämmerhirt, D., Gray, J., Venturini, T., & Meunier, A., *Advancing Sustainability Together? Citizen-Generated Data and the Sustainable Development Goals*, SSRN Electronic Journal, 2018. https://doi.org/10.2139/ssrn.3320467

Leydesdorff L. and Etzkowitz H., *The Triple Helix as a Model for Innovation Studies*, Science & Public Policy , 1998, 25(3), 195–203.

Li, W. and Quinn, P., *The European Health Data Space: An expanded right to data portability?*, Computer Law & Security Review, 2024, 52, p.105913.

Liva, Giovanni, Marina Micheli, Sven Schade, Alexander Kotsev, Matteo Gori, and Cristiano Codagnone, *City Data Ecosystems between Theory and Practice: A Qualitative Exploratory Study in Seven European Cities*, Data & Policy, 2023. https://doi.org/10.1017/dap.2023.13

LICCI - Local Indicators of Climate Change Impacts., *LICCI - Local Indicators of Climate*

Change Impacts, 2023. Retrieved May 15, 2023, from https://licci.eu/

Lightfoot, K. G., Panich, L. M., Schneider, T. D., & Gonzalez, S. L., *European colonialism and the Anthropocene: A view from the Pacific Coast of North America*, Anthropocene, 4, Elsevier, 2013, pp. 101–115. https://doi.org/10.1016/j.ancene.2013.09.002

Local Contexts., *Local Contexts*, 2023. Retrieved May 15, 2023, from https://localcontexts.org/

Lopez Solano, J., Martin, A., de Souza, S. & Taylor, L., *Governing data and artificial intelligence for all: Models for sustainable and just data governance*, European Parliament, Brussels, 2022. https://doi.org/10.2861/915401

Loukissas Y.A., *All data are local*. The MIT Press, Cambridge: MA, MIT Press, 2019.

Luers, A., Koomey, J., Masanet, E., Gaffney, O., Creutzig, F., Lavista Ferres, J. & Horvitz, E., *Will AI accelerate or delay the race to net-zero emissions?*, Nature, 628, Springer Nature, 2024, pp. 718-720. https://doi.org/10.1038/d41586-024-01137-x

Lukyanenko, R., A. Wiggins, and H.K. Rosser, 'Citizen Science: An Information Quality Research Frontier', *Information Systems Frontiers*, Vol. 22, No. 4, August 1, 2020, pp. 961–983.

Margoni, T., Ducuing, C. and Schirru, L., 'Data property, data governance and common European data spaces', *Computerrecht: Tijdschrift voor Informatica, Telecommunicatie en Recht*, (April 25, 2023). http://dx.doi.org/10.2139/ssrn.4428364

Martens, B., de Streel, A., Graef, I., Tombal, T. & Duch-Brown, N., *Business-to-Business Data Sharing: An Economic and Legal Analysis*, SSRN Electronic Journal, 2020. https://ssrn.com/abstract=3658100

Massarella, K., Krauss, J., Kiwango, W., & Fletcher, R., *Exploring Convivial Conservation in Theory and Practice: Possibilities and Challenges for a Transformative Approach to Biodiversity Conservation*, Conservation and Society, 20(2), Sage Publications, 2022, p. 59. https://doi.org/10.4103/cs.cs_53_22

Mercille J., *Inclusive Smart Cities: Beyond Voluntary Corporate Data Sharing*, Sustainability, 13(15), MDPI, 2021, pp. 1-13. https://doi.org/10.3390/su13158135

Micheli, M., *Public bodies' access to private sector data: The perspectives of twelve European local administrations*, First Monday, 27(2), University of Illinois Libraries, 2022, pp. 1-19. https://doi.org/10.5210/fm.v27i2.11720

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

Micheli, M., Ponti, M., Craglia, M., & Berti Suman, A., *Emerging models of data governance in the age of datafication*, Big Data & Society, 7(2), SAGE Publications, 2020. https://doi.org/10.1177/2053951720948087

Milan, S., 'Political Agency, Digital Traces and Bottom-up Data Practices', *International Journal of Communication*, Vol. 12, 2018, pp. 507–527.

Milan, S., and M. Gutiérrez, 'Technopolitics in the Age of Big Data', in F. Sierra Caballero and T. Gravante (eds.), *Networks, Movements & Technopolitics in Latin America: Critical Analysis and Current Challenges.*, Palgrave MacMillan, London, 2017, pp. 95–109.

Milan, S., and L. van der Velden, 'The Alternative Epistemologies of Data Activism', *Digital Culture & Society*, Vol. 2, 2016, pp. 57–74.

Monahan, T., *Monopolizing mobilities: The data politics of ride-hailing platforms in US cities*, Telematics and Informatics, 55, Elsevier, 2020, 101436. https://doi.org/10.1016/j.tele.2020.101436

Monserrate, S.G., 'The Cloud Is Material: On the Environmental Impacts of Computation and Data Storage', *MIT Case Studies in Social and Ethical Responsibilities of Computing*, Vol. Winter, 2022.

Nanda, A., Kapoor, A., Airan, A., Girish, S., Sridharan, S., Mohamed, S., & Narayan, V., *Fostering Participatory Data Stewardship: A Playbook*, Aapti institute, 2023.

Monat, J.P. and Gannon, T.F., *What is systems thinking? A review of selected literature plus recommendations*, American Journal of Systems Science, 2015, 4(1), pp. 11-26.

Montero-Pascual, J.J., Finger, M. and De Abreu Duarte, F.M., *Creating a common European mobility data space*, European University Institute, 2023.

Noto La Diega, G. & Derclaye, E., *Opening Up Big Data for Sustainability: What Role for Database Rights in the Fourth Industrial Revolution?*, in O-A Rognstad, T Pihlajarinne & J Mähönen (Eds), Promoting Sustainable Innovation and the Circular Economy: Legal and Economic Aspects (forthcoming), Routledge. http://dx.doi.org/10.2139/ssrn.4220534

OECD, Enhancing Access to and Sharing of Data: *Reconciling Risks and Benefits for Data Re-use across Societies*, OECD Publishing, 2019. https://doi.org/10.1787/276aaca8-en

OECD, *Innovative Citizen Participation and New Democratic Institutions: Catching the Deliberative Wave*. 2020, https://www.oecd.org/gov/open-government/innovative-citizen-participation-new-democratic-institutions-catching-the-deliberative-wave-highlights.pdf

Ostrom, E., *Understanding institutional diversity*, Princeton University Press, Princeton, NJ, 2005.

Ostrom, E., *Beyond Markets and States: Polycentric Governance of Complex Economic Systems*, The American Economic Review, 100(3), 2010, pp. 641–672.

Ostrom, E., Gibson, C., Shivakumar, S. and Andersson, K. *Aid, incentives, and sustainability. An institutional analysis of development cooperation*, Stockholm, SIDA, 2002.

Ottinger, G., 'Buckets of Resistance: Standards and the Effectiveness of Citizen Science', *Science, Technology, & Human Values*, Vol. 35, No. 2, March 1, 2010, pp. 244–270.

———, 'Careful Knowing as an Aspect of Environmental Justice', *Environmental Politics*, Vol. 33, No. 2, February 23, 2024, pp. 199–218.

———, 'Constructing Empowerment through Interpretations of Environmental Surveillance Data', *Surveillance & Society*, Vol. 8, No. 2, 2010.

Otto, B., ten Hompel, M. and Wrobel, S., *Designing Data Spaces: The Ecosystem Approach to Competitive Advantage,* Springer Nature, 2022, p. 580.

Parks, L., Benefit-sharing in Environmental Governance Local Experiences of a Global Concept. London, New York, Routledge, 2020.

Poikola, A., Verdonck, B., Joosten, R. editor(s), *DSSC Glossary*, Data Spaces Support Centre (DSSC), 2023.

Porciuncula, L. and de La Chapelle, B., We Need to Talk About Data Framing the Debate Around the Free Flow of Data and Data

Sovereignty, Internet and Jurisdiction policy Network, 2021.

Prainsack, B.; El-Sayed, S.; Forgó, N.; Szoszkiewicz, L. and Baumer, P., *White Paper on Data Solidarity*, Lancet And Financial Times Commission On 'Governing Health Futures 2030: Growing Up In A Digital World', 2022.

Prainsack, B. *Data Donation: How to Resist the iLeviathan*. In: Krutzinna, J., Floridi, L. (eds) The Ethics of Medical Data Donation. Philosophical Studies Series, vol 137. Springer, 2019. https://doi.org/10.1007/978-3-030-04363-6_2

Purtova, N., and van Maanen, G. (2024), 'Data as an economic good, data as a commons, and data governance', *Law, Innovation and Technology*, 16(1), 1–42. https://doi.org/10.1080/17579961.2023.2265270

Ragnedda, M., ‚Conceptualizing Digital Capital', *Telematics and Informatics*, Vo.. 35 (8) 2018, pp. 2366–2375.

Raymond, M., and L. DeNardis, 'Multistakeholderism: Anatomy of an Inchoate Global Institution', *International Theory*, Vol. 7, No. 3, 2015, pp. 572–616.

Research Data Alliance International Indigenous Data Sovereignty Interest Group., *CARE Principles for Indigenous Data Governance*, The Global Indigenous Data Alliance, 2019. Retrieved from http://gida-global.org/

Reyes-García, V., Fernández-Llamazares, Á., Aumeeruddy-Thomas, Y., Benyei, P., Bussmann, R. W., Diamond, S. K., Zorondo-Rodriguez, F., *Recognizing Indigenous peoples' and local communities' rights and agency in the post-2020 Biodiversity Agenda*, Ambio, 51(1), Springer, 2021. https://doi.org/10.1007/s13280-021-01561-7

Reyes-García, V., Tofighi-Niaki, A., Austin, B. J., Benyei, P., Danielsen, F., Fernández-Llamazares, Á., Tengö, M., *Data Sovereignty in Community-Based Environmental Monitoring: Toward Equitable Environmental Data Governance*, BioScience, 72(8), Oxford University Press, 2022, pp. 714–717. https://doi.org/10.1093/biosci/biac048

Richter, H., 'Looking at the Data Governance Act and Beyond: How to Better Integrate Data Intermediaries in the Market Order for Data Sharing', *GRUR International*, Volume 72, Issue 5, May 2023, Pages 458–470, https://doi.org/10.1093/grurint/ikad014

Robinson, C. J., Kong, T., Coates, R., Watson, I., Stokes, C., Pert, P. Chen, C., *Caring for Indigenous Data to Evaluate the Benefits of Indigenous Environmental Programs*, Environmental Management, 68, Springer, 2021. https://doi.org/10.1007/s00267-021-01485-8

Robinson, D., *Confronting Biopiracy: Challenges, Cases and International Debates*, Routledge, London, 2010.

Ryan, C., V. Salas-Wright, M. Anastario, and G. Camara, 'Making Research Matter . . . Matter to Whom?', *International Journal of Communication*, Vol. 4, 2010, pp. 845–855.

Sadowski, J., 'When data is capital: Datafication, accumulation, and extraction', *Big Data & Society*, Vol. 6(1). 2019.

Sander, I., 'What Is Critical Big Data Literacy and How Can It Be Implemented?', *Internet Policy Review*, Vol. 9, No. 2, 2020.

Sartorius, C., *Predicting Water Consumption. DAIAD Deliverable 6.3*, 2017.

Schade, S., M. Pelacho, T. (C. G.E.) van Noordwijk, K. Vohland, S. Hecker, and M. Manzoni, 'Citizen Science and Policy', in K. Vohland, A. Land-Zandstra, L. Ceccaroni, R. Lemmens, J. Perelló, M. Ponti, R. Samson, and K. Wagenknecht (eds.), *The Science of Citizen*

Science, Springer International Publishing, Cham, 2021, pp. 351–371.

Scharfbillig, M., Smillie, L., Mair, D., Sienkiewicz, M., Keimer, J., Pinho Dos Santos, R., Vinagreiro Alves, H., Vecchione, E. & Scheunemann, L., *Values and Identities - a policymaker's guide*, Publications Office of the European Union, Luxembourg, 2021, doi:10.2760/022780, JRC126150.

Schneider, G., *Framing Accountability in Business-to-Government Data Sharing: The Gap Filling Role of Businesses' Corporate Digital Responsibility*, European Business Law, 6/2022, SSRN Electronic Journal, 2022. http://dx.doi.org/10.2139/ssrn.3897263

Scholte, J.A., 'Sources of Legitimacy in Global Governance', *Outlines of Global Transformations: Politics, Economics, Law*, Vol. 12, 2019, pp. 47–76.

Schurig, T., A. Kari, and D. Fürstenau, 'THE SYMPHONY OF ORCHESTRATED PARTICIPATORY DATA SPACE GOVERNANCE: A SYSTEMATIC REVIEW', Paphos, Cyprus, 2024.

Science Europe., *Research Plays an Essential Role in Public Interest Data Use: Science Europe Response to the European Commission's Consultation on a Future Data Act*, 2021. https://www.scienceeurope.org/media/nd1jsyl3/20210827_dataact_seresponse_final.pdf

Scott, D.N., ''We Are the Monitors Now': Experiential Knowledge, Transcorporeality and Environmental Justice', Social & Legal Studies, Vol. 25, No. 3, June 1, 2016, pp. 261–287.

Second Commission Staff Working Document on Common European Data Spaces, European Commission, 2024.

Shaharudin, A., B. Van Loenen, and M. Janssen, 'Towards a Common Definition of Open Data Intermediaries', *Digital Government: Research and Practice*, Vol. 4, No. 2, June 30, 2023, pp. 1–21.

Short, Jodi L., I*n Search of the Public Interest* (June 5, 2023). Yale Journal on Regulation, Vol. 40, 2023, UC San Francisco Research Paper Forthcoming, Available at SSRN: https://ssrn.com/abstract=4470265

Shulz, S., O'Neil, M., Broca, S. & Daly, A., *Digital Commons for the Ecological Transition*, tripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society, 22(1), 2024, pp. 348-365. https://www.triple-c.at/index.php/tripleC/article/view/1456/1568

Skovgaard, L.L.; Wadmann, S., Hoeyer, K., 'A review of attitudes towards the reuse of health data among peoplein the European Union: The primacy of purpose and the common good', *Health Policy*, Vol. 123, 2019, pp. 564 – 571.

Solmaz, G., Cirillo, F., Fürst, J., Jacobs, T., Bauer, M., Kovacs, E., Santana, J.R. and Sánchez, L., *Enabling data spaces: Existing developments and challenges*. In Proceedings of the 1st International Workshop on Data Economy, December 2022, pp. 42-48.

Strandburg, K.J., Frischmann, B.M., and Madison, M.J. (Eds.), *Governing Medical Knowledge Commons*. Cambridge University Press, Cambridge, MA, 2017

Studlein. N., *Data as a Common Good: Essays on Data Portability and B2B Industrial Data Sharing*, University of Passau, 2022. https://opus4.kobv.de/opus4-uni-passau/frontdoor/index/index/docId/1120

Susha, I., *Establishing and implementing data collaborations for public good: A critical factor analysis to scale up the practice*, Information Polity, 25(1), IOS Press, 2020, pp. 3-24. https://doi.org/10.3233/IP-180117

Susha, I. & Gil-Garcia, J.R., *A collaborative governance approach to partnerships addressing public problems with private data*, Proceedings of the 52nd Hawaii International Conference on System Science, 2019. https://www.ctg.albany.edu/media/pubs/pdfs/CollaborativeGovernanceApproach.pdf

Susha, I., Janssen, M. & Velhurst, S., *Data Collaboratives as a New Frontier of Cross-Sector Partnerships in the Age of Open Data: Taxonomy Development*, Proceedings of the 50th Hawaii International Conference on System Sciences, 2017. https://aisel.aisnet.org/hicss-50/eg/open_data_in_government/4/

Susha, I., Pardo, T. A., Janssen, M., Adler, N., Verhulst, S. G., & Harbour, T., 'A Research Roadmap to Advance Data Collaboratives Practice as a Novel Research Direction'. *International Journal of Electronic Government Research* (IJEGR), 14(3), 2018, 1-11 http://doi.org/10.4018/IJEGR.2018070101

Switzer, S. & Berti Suman, A., *Private sector access to public sector personal data: exploring data value and benefit-sharing*, Scottish Government, 2023. https://www.gov.scot/publications/private-sector-access-public-sector-personal-data-exploring-data-value-benefit-sharing/

Taddeo M., 'Data philanthropy and the design of the infraethics for information societies', *Philosophical Transactions*, Vol. 28:374(2083):20160113, 2016.

Tarkowski, A. et al., *Public Data Commons: A Public interest framework for B2G data sharing in the Data Act.*, Open Future, 2022.

Taylor, L., *What is data justice? The case for connecting digital rights and freedoms globally*, Big Data & Society, 4(2), SAGE Publications, 2017, pp. 1–14. https://doi.org/10.1177/2053951717736335

Taylor, L., Floridi, L., & Van, B., *Group privacy: new challenges of data technologies*, Springer, Cham, 2017.

Tengö, M., Austin, B. J., Danielsen, F., & Fernández-Llamazares, Á., *Creating Synergies between Citizen Science and Indigenous and Local Knowledge*, BioScience, 71(5), Oxford University Press, 2021, pp. 503–518.

TEHDAS Consortium Partners, 'Report on lessons learned to be applied and recommendations 2 for data altruism practices in the implementation of construction of national and European health data spaces (including broad consent)', 2023.

Thabit, S., and Mora, L., *The collaboration dilemma in smart city projects: Time to ask the right questions*, Organization, 0(0), 2023. https://doi.org/10.1177/13505084231183949

Tombal, T., *The Rationale for Compulsory B2B Data Sharing and its Underlying Balancing Exercises*, SSRN Electronic Journal, 2021. https://ssrn.com/abstract=3984873

Van De Ven, M., A.E. Abbas, Z. Kwee, and M. De Reuver, 'Creating a Taxonomy of Business Models for Data Marketplaces', *34th Bled eConference Digital Support from Crisis to Progressive Change: Conference Proceedings*, University of Maribor Press, 2021, pp. 309–321.

Van der Sleen, T., *Voluntary business-to-government data sharing to address societal challenges: exploring actor alignment to create mutual benefits*, University of Utrecht, 2022. https://studenttheses.uu.nl/bitstream/handle/20.500.12932/43252/Thesis.pdf?sequence=1

van Rooy, A., *The Global Legitimacy Game. Civil Society, Globalization, and Protest*, Palgrave Macmillan, New York, 2004.

Veil, W., Data altruism: how the EU is screwing up a good idea, AW AlgorithmWatch gGmbH, 2021.

Vera, L.A., L. Dillon, S. Wylie, J.L. Ohayon, A. Lemelin, P. Brown, C. Sellers, D. Walker, and the Environmental Data and Governance Initiative, 'Data Resistance: A Social Movement Organizational Autoethnography of the Environmental Data and Governance Initiative', *Mobilization: An International Quarterly*, Vol. 23, No. 4, December 1, 2018, pp. 511–529.

Vera, L.A., D. Walker, M. Murphy, B. Mansfield, L.M. Siad, and J. Ogden, *When Data Justice and Environmental Justice Meet: Formulating a Response to Extractive Logic through Environmental Data Justice*, Information, communication and society, 22(7), 2019, pp. 1012–1028. https://doi.org/10.1080/1369118X.2019.1596293

Verhulst, S. G., 'The ethical imperative to identify and address data and intelligence Asymmetries', *AI & SOCIETY*, 2022.

Verstraete, M., S. Verbrugge, D. Colle, and M. Fierens, 'Assessing the Economic Viability of Data Intermediation Services as Defined in the DGA', *SSRN Electronic Journal*, 2023.

Vertesi, J. and Dourish, P., 'The Value of Data: Considering the Context of Production in Data Economies', *CSCW '11: Proceedings of the ACM 2011 conference on Computer supported cooperative work*, 2011, pp 533–542.

Viljoen, Salome., *A Relational Theory of Data Governance*, Yale LJ 131:573, 2021.

Vitale, T. (2010), 'Regulation by incentives, regulation of the incentives in urban policies', *Transnational Corporations Review*, 2(2), 35-45

Von der Leyen, U., *Political Guidelines for the Next European Commission 2024–2029*. European Commission General Publications, Strasbourg, 18 July 2024. https://commission.europa.eu/document/e6cd4328-673c-4e7a-8683-f63ffb2cf648_en

Warne, H., Dencik, L. and Hintz, A, 2021. *Advancing civic participation in algorithmic decision-making a guidebook for the public sector*. Data Justice Lab, 2021, https://orca.cardiff.ac.uk/id/eprint/143384.

Walgrave, S. and Vliegenthart, R., Protest and agenda-setting. Edited by F.R. Baumgartner, C. Breunig and E. Grossman, Comparative Policy Agendas: Theory, Tools, Data. Oxford: Oxford University Press, 2019.

Walker, M, T. Kukutai, S. Russo Carroll, and D. Rodriguez-Lonebear, eds., Indigenous Data Sovereignty and Policy, Routledge, Abingdon, 2021.

Walter, M., & Carroll, S. R., *Indigenous Data Sovereignty, governance and the link to Indigenous policy*, in M. Walter, T. Kukutai, S. R. Carroll, & D. Rodriguez-Lonebear (Eds.), Indigenous Data Sovereignty and Policy, Routledge, New York, 2021, pp. 1–20.

Waste4Think, 'Moving towards Life Cycle Thinking by Integrating Advanced Waste Management Systems', 2020. https://cordis.europa.eu/project/id/688995.

Williamson, B., Provost, S., & Price, C., *Operationalising Indigenous data sovereignty in environmental research and governance*, Environment and Planning F, SAGE Publications, 2022. https://doi.org/10.1177/26349825221125496

Woźniak-Cichuta, M., *Business-to-Business Data Sharing within the EU Digital Market (2019-2021)*, SSRN Electronic Journal, 2023. http://dx.doi.org/10.2139/ssrn.4486095

Wylie, S.A., K. Jalbert, S. Dosemagen, and M. Ratto, 'Institutions for Civic Technoscience: How

Critical Making Is Transforming Environmental Research', *The Information Society*, Vol. 30, No. 2, March 1, 2014, pp. 116–126.

Wynne, B., 'May the Sheep Safely Graze? A Reflexive View of the Expert–Lay Knowledge Divide', by S. Lash, B. Szerszynski, and B. Wynne, *Risk, Environment and Modernity: Towards a New Ecology*, SAGE Publications Ltd, London, 1998, pp. 44–83.

Zoboli, L., *Business-to-Business Data Sharing within the EU Digital Market (2019-2021): B2B Non-Personal Data Sharing in EU Legal Framework*, SSRN Electronic Journal, 2023. http://dx.doi.org/10.2139/ssrn.4462368

Zoboli, L., *Fueling the European Digital Economy: A Regulatory Assessment of B2B Data Sharing*, European Business Law Review, 31 (4), Kluwer Law International, 2020, pp. 663-692. https://kluwerlawonline.com/JournalArticle/European+Business+Law+Review/31.4/EULR2020026

LIST OF ABBREVIATIONS AND DEFINITIONS

Abbreviations	Definitions
AI	Artificial Intelligence
B2B	Business-to-Business
B2C	Business-to-Consumer
B2G	Business-to-Government
BBMRI	Biobanking and BioMolecular resources Research Infrastructure
BC	Biocultural
BDVA	Big Data Value Association
CARE	Collective Benefit, Authority to Control, Responsibility, Ethics.
CGD	Citizen-Generated Data
DCC	Data Cooperation Canvas
DGA	Data Governance Act
DIS	Data Intermediary Service
DISP	Data Intermediation Service Provider
DMP	Data management plan
DSSC	Data Spaces Support Centre
EC	European Commission
ECMWF	European Centre for Medium-Range Weather Forecasts
EDIB	European Data Innovation Board
EPOS	European Planetary Observing System
ERIC	European Research Infrastructure Consortium
EU	European Union

Abbreviations	Definitions
FAIR	Findable, Accessible, Interoperable and Reusable
GBIF	Global Biodiversity Information Facility
GDDS	Green Deal Data Space
GDPR	General Data Protection Regulation
HVD	High-Value Datasets
IDS	Indigenous Data Sovereignty
IDSA	International Data Spaces Association
IPLC	Indigenous peoples and local communities
ISO	International Organization for Standardization
JRC	Joint Research Centre
ODI	Open Data Institute
OGC	Open Geospatial Consortium
OGD	Open Government Data
PIMS	Personal Information Management Systems
RDAO	Recognized Data Altruism Organization
SDI	Spatial Data Infrastructure
SME	Small and mid-size enterprises
STS	Science and Technology Studies
TK	Traditional Knowledge
W3C	The World Wide Web Consortium
PIMS	Personal Information Management Systems

GLOSSARY

Benefit: advantages, gains, or positive outcomes that individuals or entities receive due to their actions or participation in a particular activity. Benefits can be direct or indirect and may include improvements in financial gains, enhanced reputation, and market opportunities. Unlike incentives, which are offered as inducements to influence behavior, benefits are the actual outcomes or advantages that individuals or entities experience because of their actions.

Data altruism: the voluntary sharing of data by individuals or organizations for the common good, without expecting a direct personal or financial benefit, typically to support research, public services, or societal goals.

Data ecosystem: complex socio-technical networks of people, organizations, technology and policies where various parties engage in a sustainable cycle of data sharing and value creation (Liva et al., 2023).

Data intermediation service: platform or entity that facilitates the exchange of data between different parties, through technical, legal or other means.

Data space: following the definition of the Data Spaces Support Centre (DSSC), a data space is 'an infrastructure that enables data transactions between different data ecosystem parties based on the governance framework of that data space. Data space should be generic enough to support the implementation of multiple use cases' (Poikola et al., 2023, p.5).

Common European data spaces: the common European data spaces are conceived as domain-specific data spaces where personal and non-personal data can flow within boosting growth and creating value, while respecting European rules and values and ensuring fair, secure, and trustworthy access to and use of data (EC, 2020).

Common European Green Deal data space: the common European data space that 'will interconnect currently fragmented and dispersed data from various ecosystems, both for/from the private and public sectors, to support the objectives of the European Green Deal. Also Green Deal Data Space (GDDS)' (GREAT Project, 2024).

Green Deal Data refers to environmental and environmentally relevant information, comprising both geospatial and non-geospatial data.

Data Governance: the collection of practices, processes, policies, regulations, and standards established by multiple actors or organizations to collect, share, manage and use data.

Federated Platform: technical infrastructure spanning from cloud to edge used to federate and distribute data among stakeholders.

Data Space Federated Governance: governance form characterized by 'a range of political, institutional and administrative rules, practices and (formal and informal) processes through which and how decisions are taken and implemented; decision-makers are held accountable in the development and management of [...] resources and the delivery of [...] services; and, last but not least, stakeholders articulate their interests and have their concerns considered' (Dietrick & Gutiérrez David, 2024). Federated represents the joining presence of smaller entitles, reflecting the decentralization of the data space governance, which emphasizes a 'network of decision-making across multiple levels' (Fritzenkötter et

al., 2022). These entities collaborate to ensure that they work towards agreed objectives, using agreed strategies.

Digital Twin is a digital replica of a living or non-living physical entity, a virtual representation of a connected real thing or a set of things representing a complex domain environment. It can be used to run simulations. Digital Twins have been around for decades (especially in industry), however, with the advent of transformative technologies (IoT, AI, ML, Big Data analytics, and ubiquitous connectivity) they are changing most of the society sectors –including science.

Edge computing: a methodology for optimizing cloud computing systems by performing data processing at the edge of the network, near the source of the data. For example, performing more computation at the level of the sensors capturing the data, or mobile devices like mobile phones. In this way there is less need to transfer data to centralised servers or clouds.

European single market for data: 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 also have easy access to high-quality industrial data, boosting growth and creating value (EC, 2020)

European strategy for data: launched by the European Commission, is a comprehensive framework aimed at creating a single market for data that ensures Europe's global competitiveness and data sovereignty, by facilitating data sharing and usage, fostering innovation, and ensuring trust and compliance with European regulations and values.

Governance: broadly refers to the web of actors involved, with different roles, in the process of governing a system. The term stresses a discontinuity from so-called 'command-and control' by the State, and acknowledges that a broader set of actors and institutions are also involved in managing societies like the private sector, civil society and other non-government entities.

Incentive: the rewards and punishments that are perceived by individuals to be related to their actions and those of others: 'the payments people receive or costs they have to pay, the respect they earn from others, the acquisition of new skills or knowledge are all external stimuli that may induce more of some kinds of behaviour and less of other kinds' (Ostrom et al., 2002, p.6).

Knowledge commons: establishment of community-based governance structures for the sharing, and occasionally the creation, of information, scientific findings, knowledge, data, and various other forms of intellectual and cultural resources (Strandburg et al., 2017).

Regulatory incentive: in this context it refers to legal requirements that contain incentives for encouraging environmental data access, sharing and reuse. The term 'legal requirements' refers to the obligations provided by the law.

Regulatory disincentive: in this context it refers to legal requirements that deter or discourage environmental data access, sharing and reuse.

Stakeholder: all organisations and individuals involved in, or affected by, the production, availability, and use of environmental data. Organisations can be both public and private. In the context of the GDDS, stakeholders include businesses, public sector, governments, Data Intermediary Services Providers, Data Altruism Organizations, and citizens, among the others.

Platform economy: an economy underpinned by platforms. From an economic perspective, a (digital) platform is where two or more types of users (consumers, suppliers, advertisers, software developers, etc.) come together to exchange goods, services and information. They leverage the data that they collect on user behaviour on the platform to reinforce their own position.

LIST OF BOXES

LIST OF FIGURES

LIST OF TABLES

ANNEXES

Annex 1. List of analysed entities relevant to Environmental and Geospatial Data Intermediation Services.

TABLE A1.

List of analysed data intermediaries and service providers.

ID	NAME	WEBSITE	TYPE
1	BSC EARTH SYSTEM SERVICES	HTTPS://ESS.BSC.ES/	SERVICE PROVIDER
2	Urban Big Data Centre	https://www.ubdc.ac.uk	DIS
3	Creodias	https://creodias.eu/	DIS
4	7eData	https://7edata.com/	Service provider
5	Circularize	https://www.circularise.com/	DIS
6	SyncForce	https://www.syncforce.com/	DIS
7	iPoint	https://www.ipoint-systems.com/	Service provider
8	EOMAP	https://www.eomap.com/	Service provider
9	Deltares	https://www.deltares.nl/	DIS
10	GRID Arendal	https://www.grida.no/	Service provider
11	ECMWF	https://www.ecmwf.int/	DIS
12	4sfera	https://4sfera.com/en/inici-english/	Service provider
13	space4environment	https://space4environment.com/	Service provider
14	Digi Cycle	https://www.digi-cycle.at/	Service provider
15	SpatialServices	https://www.spatial-services.com/en/spatial-services-ltd/	Service provider
16	Upstream Mobility	https://www.upstream-mobility.at/#mobilitatsplattform	DIS
17	Terrasigna	http://www.terrasigna.com/	DIS
18	HubOcean	https://www.hubocean.earth/	DISP
19	DockTech	https://www.docktech.net/	Service provider
20	GeoSystems Hellas	https://www.geosystems-hellas.gr/	Service provider
21	GHGSat	https://www.ghgsat.com/	DIS
22	GiSat	https://www.gisat.cz/	Service provider
23	Hydrologic	https://www.hydrologic.com/	Service provider
24	Mercator Ocean	https://www.mercator-ocean.eu/en/	Service provider
25	Spire	https://spire.com/	DIS
26	Elipsis Earth	https://www.ellipsis.earth/home	Service provider
27	AGRI-DataHub	https://agdatahub.eu	DISP
28	Topolytics	https://topolytics.com/	DIS

ID	NAME	WEBSITE	TYPE
29	Cervest	https://cervest.earth/	Service provider
30	Trase	https://trase.earth/	DIS
31	Mundis	https://mundiwebservices.com/	DIS
32	Catalyst	https://catalyst.earth/	Service provider
33	Eleaf	https://eleaf.com/	Service provider
34	Mitiga Solutions	https://www.mitigasolutions.com/	DIS
35	SkyBlue	https://blueskyhq.io/data-hub	DIS
36	PylonData	https://pylondata.es/	DIS
37	DATACIE	https://datacie.com	DIS
38	NORTH	https://north.io/en/	DIS
39	AINO	https://aino.world/	Service provider
40	Join Data	https://join-data.nl/en/	DISP
41	Farmdesk	https://www.farmdesk.eu/	DIS
42	Place	https://thisisplace.org/	DISP

Source: own elaboration.

TABLE A2.

List of analysed data spaces, marketplaces and providers.

ID	NAME	WEBSITE	TYPE	PHASE
1	RUDI (Rennes Metropol)	https://rudi.bzh/catalogue	Data Provider	Operational
2	GREAT project	greatproject.eu	Data space	Exploratory
3	AD4GD	https://ad4gd.eu/	Data space	Preparatory
4	GreenDealDataSpace	https://green-deal-dataspace.eu/	Data space	Preparatory
5	DS4SSCC	https://www.ds4sscc.eu/	Data space	Implementation
6	Climate data hub	https://www.eea.europa.eu/en/datahub	Data Provider	Operational
7	AGRIDATASPACE	https://agridataspace-csa.eu/	Data space	Exploratory
8	Data Space 4.0	https://digitalfactoryalliance.eu/data-space-4-0-alliance/	Data space	Unknown
9	iMagine	https://dashboard.cloud.imagine-ai.eu/marketplace	Marketplace	Operational
10	EUDAT	https://eudat.eu/	Data Provider	Operational
11	SeaDataNet	https://www.seadatanet.org/	Data Provider	Operational
12	LifeWatch ERIC	https://www.lifewatch.eu/	Data Provider	Operational
13	International Green Data Spaces (InGDS)	https://www.greendatahub.at/in-gds/?lang=en	Data space	Preparatory
14	Ocean Twin	https://ocean-twin.eu/	Digital Twin	Scaling
15	Usage	https://www.usage-project.eu/	Data Provider	Implementation
16	Waterverse	https://waterverse.eu/	Data space	Implementation
17	EarthServer	http://www.earthserver.eu/	Data space	Implementation
20	DRURAL	https://drural.eu/	Marketplace	Preparatory
21	B-CUBED	https://b-cubed.eu	Data space	Preparatory

ID	NAME	WEBSITE	TYPE	PHASE
22	ENES	https://enesdataspace.vm.fedcloud.eu/	Data space	Implementation
23	DJUST Connect	https://djustconnect.be/en	Marketplace	Operational
24	i4Trust	https://i4trust.org/experiments/cads/	Data space	Implementation
25	AgriDataSpace	https://agridataspace-csa.eu	Data space	Preparatory
26	Copernicus Data Space	https://dataspace.copernicus.eu/	Data space	Scaling
27	Destination Earth	https://destination-earth.eu/	Digital Twin	Implementation
28	Energy Web	https://www.energyweb.org/	Web3	Operational

Source: own elaboration.

TABLE A3.

List of related projects..

ID	NAME	WEBSITE
1	Biodiversity Buildings Blocks for Policy	https://b-cubed.eu/
2	ODECO	https://odeco-research.eu/
3	Cirpass	https://cirpassproject.eu/
4	EMODnet	https://emodnet.ec.europa.eu/en
5	ENVRI-FAIR	https://envri.eu/home-envri-fair/
6	EOSC Future	https://eoscfuture.eu/
7	iMagine	https://www.imagine-ai.eu/
8	interTwin	https://www.intertwin.eu/
9	Spectrum	https://www.spectrumproject.eu/
10	Decido	https://www.decido-project.eu/
11	DUET	https://www.digitalurbantwins.com/
12	Policy Cloud	https://policycloud.eu/
13	ai4publicpolicy	https://ai4publicpolicy.eu/
14	Twinergy	https://www.twinergy.eu/
15	FairCube	https://fairicube.nilu.no/
16	Green Data Hub	https://www.greendatahub.at/
17	PAIRS	https://www.pairs-projekt.de/en/energy
18	HERAKLION	https://www.heraklion-projekt.de/
19	IMPETUS	https://impetus-project.org/
20	INT:NET	https://intnet.eu/
21	Preparatory Data Space for Mobility	https://mobilitydataspace-csa.eu/
22	UPCAST	https://www.upcast-project.eu/
23	CUBE4ALL	https://incubed.esa.int/portfolio/cube4all/
24	ASPIRE	https://www.aspire2050.eu/aspire/the-association
25	Green Deal Data Observatory	https://greendeal.dataobservatory.eu/
26	Future Forests	https://futureforests.ie
27	DigiChecks	https://digichecks.eu
28	Dacapo	https://www.dacapo-project.eu

ID	NAME	WEBSITE
29	EDSC (WeTransform)	https://environmentaldataspace.com/
30	Grow Observatory	https://growobservatory.org/
31	TWIGA	https://website.twiga-h2020.eu/
32	HERMANA	https://www.hermana-colombia.co/
33	RECONECT	http://www.reconect.eu/
34	SCOREWATER	https://www.scorewater.eu/
35	G4AW	https://g4aw.spaceoffice.nl/en/
36	ODALA project	https://odalaproject.eu/
37	Eiffel	https://www.eiffel4climate.eu/
38	Open Earth Monitor	https://earthmonitor.org/
39	EO4EU	https://www.eo4eu.eu/platform
40	ENERSHARE	https://enershare.eu

Source: own elaboration.

Annex 2. National competent bodies and authorities under the Data Governance Act.

TABLE A4.

National competent bodies and authorities under the Data Governance Act. Latest update: 29/07/2024.

Member State	Competent body (Art. 7)	Competent authority for data intermediation (Art. 13)	Competent authority for data altruism (Art. 23)
Austria			
Belgium	The Federal Service Integrator, under the competence of the Federal Public Service Policy and Support	Federal Public Service of Economy, SMEs, Self-Employed and Energy	
Bulgaria	Minister of e-Government, President of the National Statistical Institute (for reuse of statistical data)	Minister of e-Government	
Croatia	Central State Office for the Development of Digital Society	Central State Office for the Development of Digital Society	
Cyprus			
Czech Republic			
Denmark	Statistics Denmark	Agency for Digital Government	
Estonia			
Finland	Statistics Finland Finnish Social and Health Data Permit Authority Findata (for secondary use of social and health care data	Finnish Transport and Communications Agency Traficom	

Member State	Competent body (Art. 7)	Competent authority for data intermediation (Art. 13)	Competent authority for data altruism (Art. 23)
France	DINUM - Direction interministérielle du numérique, placée sous l'autorité du ministre de la Transformation et de la Fonction publique	ARCEP - L'Autorité de régulation des communications électroniques, des postes et de la distribution de la presse	CNIL - La Commission Nationale de l'Informatique et des Libertés
Germany			
Greece			
Hungary	The National Data Asset Agency	The National Data Protection and Freedom of Information Authority	
Ireland			
Italy			
Latvia	Ministry of Environmental Protection and Regional Development Republic of Latvia		
Lithuania	The State Data Agency	The State Data Protection Inspectorate	
Luxembourg			
Malta			
Netherlands	Statistics Netherlands	The Authority for Consumers and Markets	
Poland			
Portugal			
Romania			
Slovakia			
Slovenia			
Spain	Deputy Directorate General for Planning and Governance of Digital Administration General Secretariat for Digital Administration. State Secre-tariat for Digitization and Arti-ficial Intelligence Ministry of Economic Affairs and Digital Transformation	Deputy Directorate General for Digital Society Directorate General for Digitization and Artifi-cial Intelligence State Secretariat for Digitization and Artificial Intelligence Ministry of Economic Affairs and Digital Trans-formation	
Sweden			

Source: https://digital-strategy.ec.europa.eu/en/library/national-competent-bodies-and-authorities-under-data-governance-act. Accessed on August 7, 2024.

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

Publications Office
of the European Union