Grant Agreement N°: 101083927
Topic: DIGITAL-2021-CLOUD-AI-01
Type of action: CSA

Green Deal Data Space and Foundation and its Community of Practice (GREAT)

D4.1: Phase 1 Governance Requirements and Endorsed Governance Scheme

Version: v1.0

Deliverable no	4.1
Work package	4
Dissemination level	PU
Due date of deliverable	M12 (31 of August 2023)
Actual submission date	31 of August 2023

Author(s)			
Partner	First Name	Last name	Email
EGI	Mark	Dietrich	Mark.dietrich@egi.eu
EGI	Marta	Gutierrez	Marta.Gutierrez@egi.eu

Contributor(s)			
Partner	First Name	Last name	Email
Seascape Belgium	Julia	Vera	julia.vera@seascapebelgium.be
ECMWF	Sebastien	Denvil	sebastien.denvil@ecmwf.int
IDC	Nevena	Raczko	nraczko@idc.com
SURF	Claudio	Cacciari	claudio.cacciari@surf.nl
SURF	Raymond	Oonk	raymond.oonk@surf.nl
IDC	Golboo	Pourabdollahian	gpourabdollahian@idc.com
UU	Kor	De Jong	k.dejong1@uu.nl
INGV	Daniele	Bailo	daniele.bailo@ingv.it
CESNET	Jiří	Sitera	sitera@cesnet.cz

EARSC	Weronika	Borejko	weronika.borejko@earsc.org

Version	Date	Released by	Comments	Document status
0.6	7 August 2023	EGI	Initial draft	Partial Draft
0.7	22 August 2023	EGI	Respond to reviewer comments, including suggestions for re-arranged content	Partial Draft
0.8	30 August 2023	EGI/IDC	Quality control	
1.0	31 August 2023	EGI	Version submitted	

Glossary

The following definitions are incorporated from the DSSC Glossary version 1.0. Where needed clarification for the purposes of the Green Deal Data Space is shown alongside the DSSC's definition.

Term	Description
Governance	GOVERNANCE is the process for making decisions about an entity: ● Choosing the questions that must be decided, such as the mission and objectives of the entity, the problems to be solved, and the needs to be addressed. ● Agreeing on the "scope" and "boundaries" of the entity, both initially and over time. ● Ensuring compliance of the entity with applicable laws and regulations. ● Deciding who should be involved in decision-making, including both the actual decision process (including activities like voting, etc.), as well as consultation about each decision. ● Decisions include those about the creation of the entity, such as its form and the relationships between and among outside parties with the entity, as well as who should participate in both decision making and governance and in the operation of the entity. ● Managing the decision-making process, recording both results and details about how these results were decided, such as who was consulted. ● Communicating about the governance process -- identifying who is involved, what decisions are being considered, what decisions have been made. ● Tracking the decisions made, monitoring compliance with these decisions, enforcing those decisions consistent with processes (which have also been decided through the governance process). ● Measuring and reviewing the performance of the entity against agreed objectives."
Data ecosystem	A collection of data and related resources, provided, produced, and/or used by a community of actors in pursuit of one or more shared objectives
Data space	(as defined by DSSC Glossary v1.0): 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.

Data space initiative	A collaborative project of a consortium or network of committed partners to deploy and maintain a data space
Digital ecosystem	A purposeful collaboration or partnership consuming, producing and providing interoperable data and related resources
Green Deal Data ecosystem	A collection of data and related resources, provided, produced, and/or used by a community of actors with the purpose of enabling the achievement of the objectives of the European Green Deal.
Green Deal Data space	(as defined by DSSC Glossary v1.0): An infrastructure that enables data transactions between different data ecosystem parties based on the governance framework of that data space. Both the data space and its related governance framework are designed to support the implementation of multiple use cases related to the European Green Deal.
Green Deal Data space initiative	A collaborative project of a consortium or network of committed partners to deploy and maintain a data space with the purpose of enabling one or more use cases associated with the European Green Deal
Green Deal Digital ecosystem	A collaboration or partnership consuming, producing and providing interoperable data and related resources with the purpose of enabling the achievement of the objectives of the European Green Deal

Abbreviations

Term	Description
AAI	Authentication and Authorisation Infrastructure
AI	Artificial Intelligence
AISBL	Association Internationale Sans But Lucratif
API	Application Programming Interface
BDI	Basic Data Infrastructure
BDVA	Big Data Value Association
CC-BY-NC	Creative Commons Attributions Non-Commercial
CDO	Chief Data Officer
CINEA	Climate, Infrastructure and Environment Executive Agency
CIO	Chief Information Officer
COBIT	Control Objectives for Information and Related Technologies
CoP	Community Of Practice
COP	Conference Of Parties
CSA	Coordination and Support Action
D-MRV	Digital Monitoring Reporting and Verification
DA	Data Act
DAltOs	Data Altruism Organisations
DE	Digital Ecosystem
DEP	Digital Europe Programme

DevOps	Development and Operations
DG	Data Governance
DG-MARE	Directorate-General for Maritime Affairs and Fisheries
DGA	Data Governance Act
DMA	Digital Markets Act
DPSIR	Drivers, Pressures, State, Impact, and Responses
DPV	Data Privacy Vocabulary
DS	Data Space
DSA	Digital Services Act
DSBA	Data Space Business Alliance
DSG	Data Space Governance
DSSC	Data Space Support Centre
EC	European Commission
ECJ	European Court of Justice
ECMWF	European Centre for Medium Range Weather Forecast
ECO	Executive Coordination Office
EDIB	European Data Innovation Board
EDIC	European Digital Infrastructure Consortium
EG	Expert Group
EGD	European Green Deal
EGDS	European Green Deal Data Space
EMFAF	European Maritime, Fisheries and Aquaculture Fund

EMODnet	European Marine Observation and Data Network
EPOS	European Plate Observing System
ERIC	European Research Infrastructure Consortium
ESD	European Strategy for Data
ESD	European Strategy for Data
EUH4D	European Federation of Data Driven Innovation Hubs
F-L	Formation Legal
FAIR	Findability, Accessibility, Interoperability, and Reusability
FP	Focal Points
FRAND	Fair, Reasonable and Non-Discriminatory
GD	Green Deal
GDDI	Green Deal Data Space Initiative
GDDS	Green Deal Data Space
GDPR	General Data Protection Regulation
GEO	Group on Earth Observation
GGF	Generic Governance Framework
GOS4M	Global Observation System for Mercury
GREAT	Green Deal Data Space Foundations and Community of Practice
HVD	High Value Datasets
ICS-C	Integrated Core Service Central hub
IDS	International Data Spaces
IDSA	International Data Spaces Association

IGPMM	Information Governance Process Maturity Model
IMEC	International Maritime Employers' Council
INSPIRE	Infrastructure for Spatial Information in Europe
IPR	Intellectual Property Rights
IR	Implementing Rules
ISMS	Information Security Management System
ISO/IEC	International Organisation for Standardisation/International Electrotechnical Commission
IT	Information Technology
JRC	Joint Research Centre
KPI	Key Performance Indicators
M-C	Monitoring-Compliance
M-I	Monitoring-Improvement
MCM	Minamata Convention on Mercury
MKEG	Marine Knowledge Expert Group
ML	Machine Learning
MS	Member State
NIS	Network and Information Security
O&M	Operations Monitoring
OAuth2	Open Authorisation 2
ODRL	Open Digital Rights Language
OECD	Organisation for Economic Co-operation and Development
OGC	Open Geospatial Consortium

P2B	Platform-to-Business
PIMS	Privacy Information Management System
PSD2	Payment Services Directive 2
SAB	Scientific Advisory Board
SC	Steering Committee
SCA	Strong Customer Authentication
SDG	Sustainable Development Goals
SDK	Software Development Kit
SE	Social Enterprises
SME	Small Medium Enterprise
TCS	Thematic Core Services
UN	United Nations
US	United States
UU	Utrecht University
VLIZ	Flanders Marine Institute
WIS	WMO Information System
WMO	World Meteorological Organisation
XACML	Extensible Access Control Markup Language

Executive Summary

Europe has taken bold action to address the environmental and societal challenges of our times. The European Green Deal stands as a key priority of the European Commission, launching a set of ambitious strategic actions with the goal of achieving climate neutrality and net-zero emissions by 2050.To realise these targets, the establishment of a single data market where data flows seamlessly across sectors and borders in a sovereign manner, is crucial. The European Strategy for Data introduces the launch of the sectoral data spaces and lays a legislative framework for trusted data sharing. The Green Deal Dataspace (GDDS), a cross-sectoral dataspace, touches all economic sectors of society. The challenges associated with designing a dataspace that spans all sectors, involving multiple stakeholders from diverse scopes, are complex and unprecedented. Given the global nature of the Green Deal, the data space needs to accommodate global, regional, national and European initiatives. The governance of data space will be guided by European values, ensuring maximum reuse of data while duly respecting data sovereignty.

The GREAT project is building a set of pillars to support an implementation roadmap for the GDDS, namely a reference blueprint architecture a governance framework and an inventory of high priority datasets. These pillars are based on the needs of a Community of Practice that involves contributors aligned with the goals of the European Green Deal.

This document presents the outcomes of Phase 1 of the project, outlining a preliminary proposal for the governance framework of the GDDS. The activities that have led to this framework include an assessment of the current landscape on dataspaces initiatives, requirements gathering from the reference use cases and data sharing initiatives, consultations with stakeholders, advisory board and ethical advisor, validation of results and alignment with the Data Space Support Centre (DSSC) assets and other sectoral data spaces.

The governance of digital platforms, based on existing literature, forms a solid foundational basis of the GDDS governance framework. This foundation is enhanced with the all the emerging activity from the data spaces community and well-established data management practices. The document proposes a generic governance framework that integrates insights from both digital platforms and dataspace initiatives. This generic framework is then adapted to identify the governance requirements specific to the Green Deal Data Space, incorporating insights from reference use cases, data sharing initiatives, and stakeholders.

These requirements will undergo further refinement in Phase 2 of the project, focusing on cross-disciplinary aspects of data sharing within the Green Deal Data Space and across sectoral data spaces.

Numerous established data sharing initiatives within dedicated domains represent important stakeholders for, and will become key participants in, the GDDS. Their significant investments of resources, effort and time must be respected. The governance of the GDDS builds on best practices from these initiatives evaluating how these communities can contribute to the data space based on their needs, value proposition and contributions toward the overarching GDDS objectives.

The evolution of the DSSC assets, the growing suite of horizontal data regulation and the emergence of common middleware for common European Data Spaces present an evolving ecosystem that highlights needs for a flexible and dynamic approach to governance. Numerous critical questions highlighted throughout the document (as GUIDANCE requirements) remain unanswered today, and the GREAT project needs these answers in order to provide a clear pathway, and a roadmap forward.

Table of Contents

1. Introduction

The "Green Deal Data Space Foundations and Community of Practice" project ("GREAT" Project) is a coordination and support action (CSA) funded by the Digital Europe program of the European Commission (EC), preparing for the implementation of the common pan-European data space related to the European Green Deal, namely the Green Deal Data Space (GDDS). The creation of the GDDS was contemplated in the European Strategy for Data.

1.1. Background

1.1.1. The European Strategy for Data

Data-driven innovation plays a key role in the digital transformation of our society and organisations[1]. The priority "A Europe fit for the digital age"[2] guides the European Commission's policy agenda for the period of 2019-2024, culminating in the EC's vision for Europe's digital transformation "2030 Digital Compass: the European way for the Digital Decade"[3] which sets ambitious targets aimed at strengthening digital sovereignty through specific actions on data, technology and infrastructures. The Annual Single Market Report[4], published in 2023, marks the 30th anniversary of the Single Market, and highlights the ambition to create a single EU data economy through a data-driven Single Market where interoperability within and across data spaces is ensured.

In February 2020, the European Commission (EC) published a Communication introducing "A European strategy for data"[5] (ESD) for the creation of "a single European data space – a genuine single market for data, open to data from across the world". The strategy to achieve this vision is structured around four main pillars:

- A cross-sectoral governance framework for data access and use;
- Enablers: Investments in data and strengthening Europe's capabilities and infrastructures for hosting, processing, and using data, interoperability;
- Competences: Empowering individuals, investing in skills and in SMEs;
- Common European data spaces in strategic sectors and domains of public interest.

According to the ESD, Data Spaces should foster an ecosystem (of companies, civil society and individuals) that will facilitate the creation of new products and services, based on more accessible data. In addition, what distinguishes the Common European Data Spaces from other data sharing initiatives is its focus on preserving European values, balancing the flow and wide use of data, while preserving high privacy, security, safety and ethical standards. One of the nine proposed

[1] Granell, C., Mooney, P., Jirka, S., Rieke, M., Ostermann, F., Van Den Broecke, J., Sarretta, A., Verhulst, S., Dencik, L., Oost, H., Micheli, M., Minghini, M., Kotsev, A. and Schade, S., Emerging approaches for data-driven innovation in Europe: Sandbox experiments on the governance of data and technology, EUR 30969 EN, Publications Office of the European Union, Luxembourg, 2022, doi:10.2760/511775.

[2] European Commission, Directorate-General for Communications Networks, and Content and Technology, *Shaping Europe's digital future*, Publications Office, 2020, https://data.europa.eu/doi/10.2759/091014

[3] European Commission, Directorate-General for Communications Networks, Content and Technology, 2030 digital compass – The European way for the digital decade, Publications Office, 2021, https://data.europa.eu/doi/10.2759/425691

[4] European Commission, Commission Staff Working Document 2023 Annual Single Market Report: Single Market at 30, SWD(2023) 26 final, 2023, https://op.europa.eu/s/yXTN

[5] European Commission, "A European strategy for data." COM(2020) 66 final 2020. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020DC0066&from=EN

common European data spaces is the GDDS. The GREAT project is charged with developing an implementation roadmap for the GDDS, including a governance scheme, technical blueprint, and priority datasets.

To support the ESD, in November 2020, the EC proposed a Data Governance Act[6] aiming at increasing trust in data sharing and facilitating data reuse. In February 2022, the EC proposed a Data Act[7] to make more data available for use in line with EU rules and values. The Data Governance Act creates the processes and structures to facilitate data exchange, while the Data Act clarifies who can create value from data and under which conditions. Finally, in the framework of the Open Data Directive[8], the European Commission adopted in December 2022 an Implementing Act[9] focused on high value datasets, which provide important benefits for society, the environment and the economy. Those "High Value Datasets" (HVDs) will have to be made available free of charge, in machine-readable format, by public sector organisations.

1.1.2. The European Green Deal

In parallel with the "digital transition" described above, there is an equally important "green transition." The European Commission demonstrated unprecedented leadership in December 2019 when it unveiled its flagship action plan[10] to tackle climate change, the European Green Deal. Through this strategy, the European Union (EU) aims to become the first resource-efficient and competitive economy without net emissions of greenhouse gases by 2050.

The European Green Deal charts a comprehensive course for action, supported by a growing number (now over 149[11]) of legislative and regulatory actions. The Green Deal sets ambitious objectives across a number of priority areas of action, including restoring degraded ecosystems at land and sea across Europe with the 2030 Biodiversity Strategy[12] and reducing greenhouse gas emissions to zero by 2050 with the European Climate Law[13] and the Zero Pollution Action Plan[14].

---

[6] European Commission, REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act) https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52020PC0767

[7] European Commission, "Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on harmonised rules on fair access to and use of data (Data Act)." COM(2022)68 final https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52022PC0068&from=EN

[8] 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 (recast) (OJ L 172 26.06.2019, p. 56, ELI: http://data.europa.eu/eli/dir/2019/1024/oj)

[9] 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." 2023. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32023R0138&from=EN

[10] European Commission, "COMMUNICATION FROM THE COMMISSION The European Green Deal" COM(2019) 640 final https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52019DC0640

[11] https://www.europarl.europa.eu/legislative-train/theme-a-european-green-deal

[12] European Parliament, Directorate-General for Internal Policies of the Union, Nègre, F., The EU 2030 biodiversity strategy, European Parliament, 2020, https://data.europa.eu/doi/10.2861/545892

[13] Regulation (EU) 2021/1119 of the European Parliament and of the Council of 30 June 2021 establishing the framework for achieving climate neutrality and amending Regulations (EC) No 401/2009 and (EU) 2018/1999 ('European Climate Law') (OJ L 243 09.07.2021, p. 1, ELI: http://data.europa.eu/eli/reg/2021/1119/oj)

[14] European Commission, COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS Pathway to a Healthy Planet for All EU Action Plan: 'Towards Zero Pollution for Air, Water and Soil' COM(2021)400 final https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021DC0400&from=EN

In addition to regional action, part of the action plan is to increase the EU's "green diplomacy" and demonstrate EU leadership in multilateral fora to increase collective effort and reach the objectives of the Paris Agreement[15] and the United Nations Sustainable Development Goals (UN SDGs).

Ambitious action plans like the European Green Deal require an abundance of resources, including viable data. Data allows responsible stakeholders, including governments at multiple levels, to identify risks, tailor policy response and resource allocation, monitor progress and identify trends. However, serious data gaps remain in the global fight against climate change and environment-related risks. (According to the UNEP report Measuring Progress Towards Achieving the Environmental Dimension of the SDGs, "there is too little data to formally assess the status of 63 of the 93 environment-related SDGs indicators" [16].) Since many consequences of climate change are irreversible, data gaps and analytics deficits need to be addressed.

1.1.3. The Green Deal Data Space

The Green Deal Data Space (GDDS) stands at the intersection of these two major European policy initiatives: the EU Strategy for Data and the European Green Deal. The GDDS will be designed and implemented to exploit the potential of data to effectively support the Green Deal priority actions, empowering policy makers, businesses, researchers and citizens, from Europe and around the world, to jointly tackle issues such as climate change, circular economy, zero pollution, biodiversity protection or deforestation, including providing assurance of compliance with policies and regulations.

Out of the many European Green Deal strategic actions, the GREAT project focuses on three priorities – Biodiversity 2030; Zero Pollution; and Climate change – to effectively capture the diversity of requirements across the full range of European Green Deal initiatives. These three actions are interlinked with other EGD strategic actions and approximate the full scope of the GDDS, as well as complementing actions that are also being addressed by other thematic data spaces (such as the "Farm to Fork Strategy", which is also addressed by the common European agricultural data space[17]).

1.2. Methodology and Structure of This Deliverable

This report develops the proposed Green Deal Data Space Governance Framework by creating a generic governance framework (GGF), reflecting best practice, and then adapting and refining that generic framework to the specific needs of the European Green Deal as well as future EU environmental strategies, policies and action plans. In particular, the Green Deal Data Space Governance Framework should support the agreed mission, vision and objectives of the Green Deal Data Space as they evolve over time.

To orient the discussion of governance, Chapter 2 explores the context of data spaces, addressing general questions of objectives and purpose, definitions, scope and relationships with other initiatives.

[15] https://unfccc.int/process-and-meetings/the-paris-agreement

[16] United Nations Environment Programme (2021). Measuring Progress: Environment and the SDGs. Nairobi.https://www.unep.org/resources/report/measuring-progress-towards-achieving-environmentaldimension-sdgs

[17] https://digital-strategy.ec.europa.eu/en/library/common-european-data-spaces-agriculture-and-mobility

Chapter 3 focusses on governance itself, providing a pragmatic definition, and identifying two dimensions of governance to help organise the analysis:

- 4 layers of governance: Legal and regulatory; Digital ecosystem & digital platform; Data space; and Data, and
- 4 lifecycle stages of governance: Formation; Operation; Monitoring; and Sustainability).

Chapter 4 presents a Generic Governance Framework that consolidates learnings from literature on both digital platforms as well as on a range of "data sharing" initiatives and projects. Although termed a "generic" framework, in fact our analysis incorporates relevant requirements from EU legislation and regulation. Generic governance requirements are called-out and numbered for subsequent tracking through implementation.

Using the framework presented in Chapter 4, Chapter 5 considers the specific governance needs and requirements of the European Green Deal community of practice, particularly those identified by the projects' use cases and related Task Forces, as well as more general requirements identified by specific community members. Chapter 6 presents existing governance structures for several key data sharing initiatives examined by the GREAT Project, providing inspiration for models that might be used for the GDDS itself.

WP4 has been working in Phase 1 of the GREAT Project to gather the specific needs of the European Green Deal community of practice. The effort in Phase 1 has helped to direct the overall approach presented in Chapter 4, and preliminary observations are collected in Chapters 5 and 6. This effort will continue and expand in Phase 2, expanding the range of use cases, problems and needs to be considered, and examining the generic framework presented here to identify required changes and additions.

2. Project Context

Even before we consider the problems to be addressed by a Green Deal Data Space, it is useful to put data spaces into context – how are they defined, what are they supposed to do, how do they relate to existing data, services and even data sharing initiatives?

2.1.Definitions of Data Spaces – Implications for Their Design

Several definitions have been proposed for Data Spaces, including:

- A recent Digital Europe call for proposals refers to a data space as "data infrastructure with tailored governance mechanisms that will enable secure and cross-border access to key datasets in the targeted thematic area"[18] .
- The Data Spaces Support Centre (DSSC) initially defined a data space as a "decentralised, governed and standard-based structure to enable trustworthy data sharing between the data space participants on a voluntary basis"[19].
- The Data Governance Act (DGA) and the Data Act (DA) define a data space as a "purpose- or sector-specific or cross-sectoral interoperable frameworks of common standards and

[18] DIGITAL, Call for proposals: Cloud Data and TEF (DIGITAL-2022-CLOUD-AI-02), Version 1.0, 2022, https://ec.europa.eu/info/funding-tenders/opportunities/ docs/2021-2027/digital/wp-call/2022/ call-fiche digital-2022-cloud-ai-02 en.pdf

[19] Data Spaces Support Centre, 2022 www.dssc.eu

practices to share or jointly process data for, *inter alia*, the development of new products and services, scientific research or civil society initiatives"[20] [21].

- In version 1.0 of its Glossary[22], the DSSC now defines a "data space" as "an infrastructure that enables data transactions between different data ecosystem parties based on the governance framework of that data space".

While most of these definitions focus on enabling data sharing alone, the EU's DGA and DA definitions harness such data transactions for "the development of new products and services, scientific research or civil society initiatives" – highlighting that data spaces are expected to have greater impact than simply enabling data transactions. The GREAT project has chosen to work with the DSSC's latest definition of data spaces, since it identifies the governance framework not only as a characteristic of a data space, but in some sense as the definition of that data space. It also highlights the possibility that different governance frameworks may be needed for different sets of stakeholders and use cases. GREAT will also consider what is required to enable the "higher level" objectives contemplated by the DGA and proposed DA.

2.2.How are data spaces different from existing data sharing initiatives?

Today there are numerous data sharing initiatives in various sectors of the economy. Most of these efforts focus on data from a limited range of sources, with similar characteristics or attributes.

- Data sharing initiatives for research data are the most visible of these efforts, but in general they concentrate on open data and handle sensitive or confidential data only on an exception basis, rather than through embedded access and use control mechanisms.
- Over the last few years, initiatives such as International Data Spaces Association (IDSA)[23] and Gaia-X[24] have worked with the business community to create frameworks and tools to allow business and industry to share confidential data with their business partners, while maintaining control ("sovereignty") over how that data would be accessed and used.
- Sharing sensitive data, particularly personal health data, has been the subject of recent efforts such as the "Joint Action Towards the European Health Data Space – TEHDAS"[25] and the "Healthy Cloud" project[26], as well as the European Health Data Space[27] now in development.

Table 1 attempts to summarise the key attributes of data sharing for these three categories of data. Existing data sharing initiatives are mostly designed to accommodate the attributes of data sharing specific to one category of data.

[20] Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance (Data Governance Act), Recital 27. http://data.europa.eu/eli/reg/2022/868/oj
[21] https://www.europarl.europa.eu/doceo/document/A-9-2023-0031EN.html
[22] https://dssc.eu/space/Glossary/55443460/DSSC+Glossary+?attachment=/rest/api/content/55443460/child/attachment/att110362680/download&type=application/pdf&filename=DSSC-Data-Spaces-Glossary-v1.0.pdf
[23] https://internationaldataspaces.org/
[24] https://gaia-x.eu/
[25] https://tehdas.eu/
[26] https://healthycloud.eu/
[27] https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en

Table 1: Data Sharing Attributes of Major Categories of Data[28]

Attributes	Categories of Data		
	Research, Public Data	Sensitive Data	Industry Confidential
Security	Not specified	✓	✓
Known Parties	Anonymous Access OK	Strong assurance	Strong assurance
Visibility	Open; 1:Many or 1:Any	Sovereignty, GDPR; 1:1 or 1:Few	Sovereignty; 1:1 or 1:Few Data Act → 1:Few or 1:Many
Findability	✓	Five Safes[29]	Sovereignty, tempered by Data Act
Accessibility	✓		
Interoperability	increasing		
(Re)usability	✓		
Quality Fit for Purpose	Peer Review	Ethics Review, GDPR	Opportunistic, tempered by AI Act?
Purpose Objectives	Advancement of Knowledge	MUST be defined up front	Solve my problem, Competitive advantage

The challenge for most common European data spaces is to design mechanisms for data sharing that will accommodate different categories of data in a robust and scalable manner.

2.3.Objectives of Data Spaces – How do they create value?

The objectives to be achieved by a Data Space are key to its design and governance. To structure this dimension, the GREAT project has developed a taxonomy of possible data space objectives, illustrating the potential cumulative value they might create (see Table 2):

Table 2: Taxonomy of Data Space Objectives

Objective Level	Description
Level 0: Presence of Many Parties, Relevant Parties	A well identified Community of Practice with Participants that have a good understanding of their role and commitment towards the data space is in place.

[28] Developed by the authors.

[29] Tanvi Desai, Felix Ritchie and Richard Welpton. Five Safes: designing data access for research. University of the West of England, Economics Working Paper Series no. 1601. 2016. https://www2.uwe.ac.uk/faculties/BBS/Documents/1601.pdf

Level 1: **Level 0 + Broad Information resource**	Relevant data and services from possibly diverse sources are available with easy search, browse, access, use, consistent metadata and interoperable with each other.
Level 2: **Level 1 + Quality**	Data is labelled to specify the quality processes it has been subject to, which may include indicators such as accuracy, precision, defined procedures, mechanisms for review, errata and retraction, spatio-temporal consistency and sustainability or reliability of the data in the future and accessibility over time.
Level 3: **Level 2 + Analysis**	Various analytical tools are available, not just to transform grids, subset or visualise on individual datasets, but to bring different data across domains together to allow insights, enabling data integration and data fusion capabilities. Quality information is incorporated into the resulting product(s) so that analytical results have their own quality indicators.
Level 4: **Level 3 + Actionable Insights**	Analysis, or even data without analysis, can be targeted to a user's needs (e.g., "give me data as well as forecasts and risk assessment about my farm, about all my corporate locations, about my house"). This can include alerts if the situation changes, or new data shows a new trend.
Level 5: **Level 4 + Aggregation/ Analysis of impact**	Risks can be aggregated across sectors, jurisdictions, etc.; impact of actions taken in the past can be analysed, impact of current actions can be modelled. Overall assessments are updated as new data arrives.
Level 6: **Level 5 + Performance Monitoring**	Forecast impacts of various actions can be developed, and then new observations can be compared against the forecast.
Level 7: **Level 6 + Target Setting**	To support some use cases, particularly policy development use cases, different scenarios need to be modelled, forecasts produced, and then performance assessed against targets. As new data arrives, forecasts are updated, target status is updated and alerted

As noted above, the higher levels of this taxonomy align with the DGA's and DA's vision of data spaces enabling more than just data sharing and data transactions. The analyses, insights, etc. created at each level could be packaged as "public good" outputs or as services or products delivered or "resold" by new or existing businesses. Participants, including data and service providers as well as consumers of those resources, can participate in specific use cases supported by the data space, targeting objectives at different levels and sharing in the value created through a variety of business models.

This taxonomy aligns with the European Environmental Agency's DPSIR[30] framework (looking into **D**rivers, **P**ressures, **S**tate, Impact, and **R**esponses): The DPSIR framework highlights the need for clear and specific information on several factors in an interlinked socio-economic and ecological system. It helps define what is known about:

- **D**riving forces; and their resulting environmental
- **P**ressures; on environmental and socio-economic
- **S**tates, including the
- **I**mpact resulting from these pressures, and
- the subsequent societal **R**esponses.

Driving forces could be any kind of human activity causing environmental degradation. The results are pressures such as emissions or waste. These in turn alter the environmental state (physical, chemical and biological) and 'impacts' on ecosystems, human health and functions, eventually leading to political 'responses' (prioritisation, target setting, laws).

This taxonomy also supports the range of functions required for e.g., the digital monitoring, reporting and verification (D-MRV) system[31] contemplated to underpin future carbon markets under the goals of the Paris Agreement.

2.4.Stakeholder Community, Digital Ecosystem and Related Data Spaces

An analysis of data spaces by the EC Joint Research Centre[32] concludes: "*There is **no single technical or organisational approach** [emphasised by the JRC] that can be applied for the establishment of common European data spaces. […] Therefore, a community-based approach through co-creation and co-design of data spaces that considers the domain-specific context is the only feasible way forward that would ensure buy-in by a broad spectrum of stakeholders*." This community-based approach is promoted by the DSSC, as well as several other initiatives in the data space community (e.g., SITRA[33] and the Data Sharing Coalition[34]).

The DSSC, SITRA and Data Sharing Coalition each describe a process by which a community of stakeholders starts to collaborate through the sharing and exchange of data and other digital resources, to reach objectives (which could be mapped to the taxonomy presented in Table 2) that would be impossible without such collaboration. This collaboration has been termed a "data ecosystem"[35] by the DSSC and others. We expand the scope of this term slightly, to "digital ecosystem", which we propose to define as a "*purposeful collaboration or partnership consuming, producing and providing interoperable data and related resources*". A Digital Ecosystem is organised around its purpose and refers to both the community of practice pursuing this purpose as well as

[30] https://www.eea.europa.eu/help/glossary/eea-glossary/dpsir

[31] https://pmiclimate.org/publication/digital-monitoring-reporting-and-verification-systems-and-their-application-future

[32] Farrell, E.; Minghini, M.;Kotsev, A.; Soler-Garrido, J.; Tapsall, B.; Micheli, M.; Posada, M.; Signorelli, S.; Tartaro, A.; Bernal, J.; Vespe, M.; Di Leo, M.; Carballa-Smichowski, B.; Smith, R.; Schade, S.; Pogorzelska K.; Gabrielli, L.; De Marchi, D., *European Data Spaces: Scientific insights into data sharing and utilisation at scale*, Publications Office of the European Union, Luxembourg, 2023, doi:10.2760/400188, JRC1

[33] SITRA, Rulebook for a Fair Data Economy, 2022, https://www.sitra.fi/en/publications/rulebook-for-a-fair-data-economy

[34] https://datasharingcoalition.eu/our-approach-and-tools/

[35] Otto, B., Lis, D., Jürjens, J. et al. (2019). Data Ecosystems. Conceptual Foundations, Constituents and Recommendations for Action. Fraunhofer Institute for Software and Systems Engineering ISST

the collection of data, tools, services, and other resources held or controlled by the community of practice and employed in these collaborations. A Digital Ecosystem contains a corresponding Data Ecosystem. A Data Ecosystem excludes the IT infrastructure for storing and processing the data that is part of the Digital Ecosystem. This is a fine distinction, but it recognizes that this IT infrastructure was established to meet a range of other objectives separate from the objectives of the various data spaces and data space initiatives that will need to work with this infrastructure. Figure 1 illustrates these three concepts.

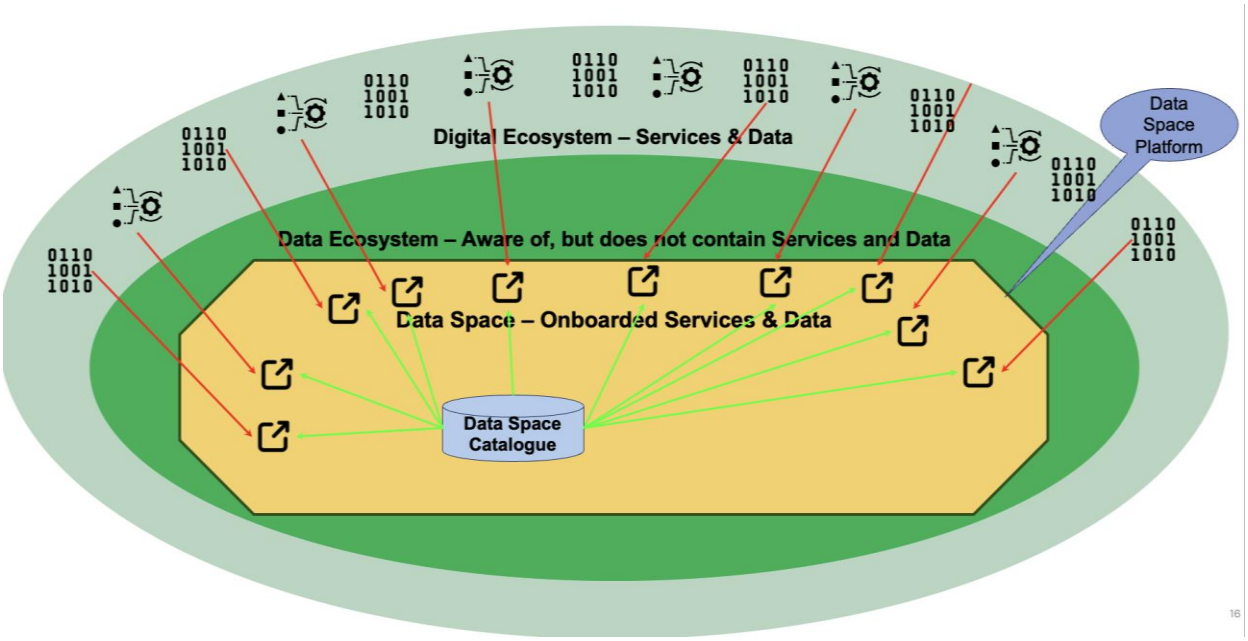

Figure 1: Relationship between Digital Ecosystem, Data Ecosystem, and Data Space

.

2.5.Challenges to Creating a Single Data Space: Data Space Initiatives as an Evolutionary Step

Each Digital Ecosystem and its related Data Ecosystem share the collective purposes of the related community of practice. The communities of practice for the Green Deal and for several other sectors are quite diverse, as are the collective purposes in each sector and the related sets of data and services needed to achieve those objectives. In most sectors there are already a number of data sharing initiatives, sometimes working toward similar goals in different places (e.g. multi-model mobility in different urban areas, environmental monitoring in different countries), and sometimes using similar methodologies to solve different problems (e.g. big data being used for precision agriculture, or for traffic monitoring).

Most data sharing initiatives begin their lives attempting to address one or a small number of use cases, bringing together data providers, data users and service providers to achieve well defined objectives. Data is not provided for general use, but for specific uses by specific users, typically in the context of informal agreements. The number of participants is limited, so trust is easier to create because participants are well known – in many cases through previous personal or organisational relationships.

The promise of data spaces is that data that has already been provided to such data sharing initiatives, addressing a select number of use cases, might also be useful for other use cases. It is hoped that the marginal "cost" of providing this data for another purpose should be low or zero, all things being equal. However, all things are not equal, because as we evolve from limited data sharing initiatives to larger initiatives, personal relationships and informal agreements need to be replaced by formal trust frameworks and explicit terms and conditions.

As an evolutionary step between a data sharing initiative and a formal data space, we exploit the concept of a "data space initiative" defined by the DSSC as "a collaborative project of a consortium or network of committed partners to deploy and maintain a data space". Such data space initiatives can start with existing data sharing initiatives, with current stakeholders, data and services, and consider how they might formalise their relationships as a sort of "mini data space", as well as considering their willingness and ability to support expanded or new use cases.

Section 5.2 explores the diversity of the Green Deal community of practice and their possible objectives and supported use cases.

2.6.Where is the Data? How is Data Actually Accessed?

None of the definitions of Data Spaces provided in Section 2.2 specify where the data itself is stored, or how it can be accessed. As an important contextual factor, the GREAT project considers the data itself to be stored inside the scope of the Digital Ecosystem, but outside any Data Space or Data Space Initiative defined within it, while metadata and well-defined interfaces "inside" the data space initiatives enable access to both data and services (several exceptions to this premise are detailed below). Figure 4 illustrates this concept.

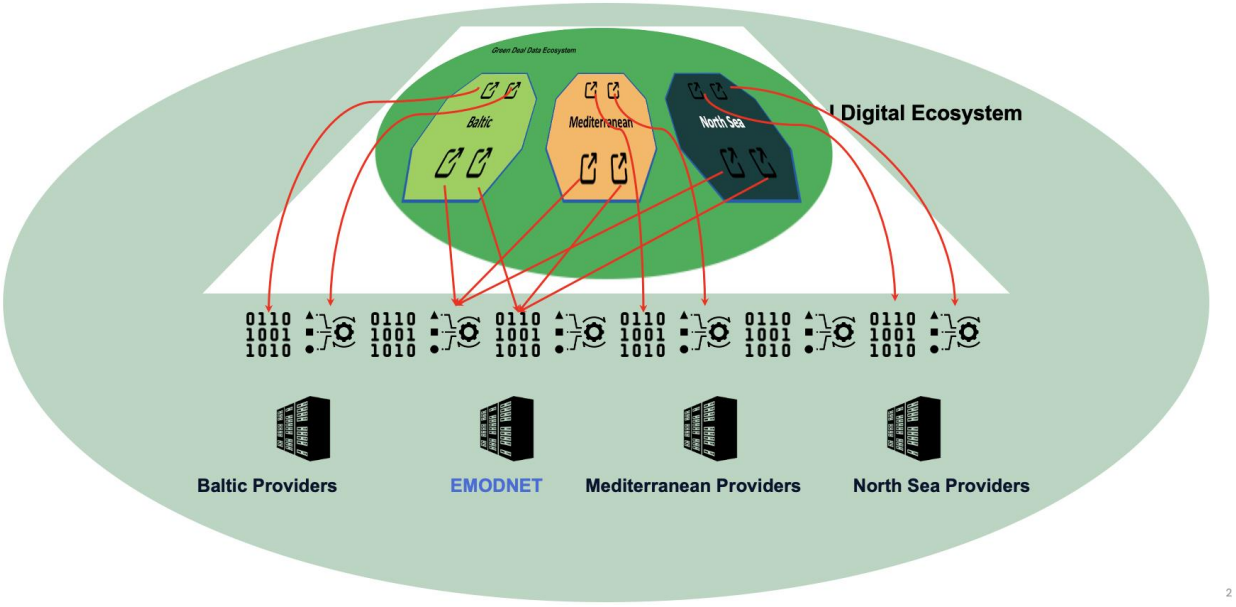

Figure 2: Services and Data are Inside the Digital Ecosystem, but Outside the Data Ecosystem

For "public data", such as data listed at EMODNet (https://emodnet.ec.europa.eu/en), web links point to online files that can be anonymously downloaded by visitors from the website, or through

API calls that do not require user authentication[36], in both cases pointing to IT systems operated by the data holder. For a given item of data, the same links can be included in one or more data spaces to enable access and use.

By contrast, for non-public data, data providers need to specify the criteria for data access as well as data use. These criteria often require identification of the data user/consumer as well as a range of other information needed by the provider to assess the request for access. Once access (and use) are approved, the data provider would enable access through one or more methods:

- Providing a URL for download to the user's own infrastructure (a local computer, or through a data transfer protocol to IT infrastructure with more storage capacity and/or computational power).
- Enabling access to a service to process requests for data (enabling human access through a web browser or other interface, and/or machine access via APIs, and possibly supporting search queries to filter the results).
- In some cases, data is structured as "streaming" data, and an authorised user can designate a location where the streaming data would be sent.
- In other cases, the data provider may also provide the user with access to local IT capabilities to support local processing of the data. This "compute-to-data" or "data visiting" scenario is often required when the data are too sensitive – or too large – to be transferred to another location or facility.

The data space community (DSSC, Data Spaces Business Alliance, Gaia-X, BDVA, IDSA, FIWARE, etc.) distinguishes between access and use policies, since the principle of data sovereignty requires that data holders and data rights holders should have the ability to control how their data is used. GREAT proposes a governance approach that requires data holders to clearly define both access and use policies. Since automatic control over data use (as opposed to data access) remains a key challenge, GREAT also considers the possibility that usage control might be achieved through explicit governance mechanisms, such as legal agreements, combined with a robust trust framework.

To reiterate, data are not by definition "stored", nor are related services "operated", by a data space or data space initiative, but they are part of the surrounding digital ecosystem, and they are potentially accessible from one or more data spaces.

We separate data from its storage and processing in order to make it easier to define what is needed to implement data spaces. However, there are important dependencies between the data and its physical storage and processing that cannot be overlooked and that may need to be addressed in parallel with the creation of any data space, as follows:

- Providing a repository for valuable data, which would otherwise be lost, could be an important activity for the Digital Ecosystem, for several reasons:
    - Given the volume of data that may be important for reaching the Stakeholder Community's objectives, being able to store this "big data" can be a challenge. For example, the Destination Earth Data Lake will allow data outputs from the project

[36] E.g. https://data.europa.eu/api/hub/search/#tag/Search

to be stored for only a limited period. Longer term storage of important results may be needed.
    - Large data outputs generated by publicly funded research projects typically cannot be stored after funding for the project ends.
    - Citizen generated data is recognized as a data source of growing importance for many Data Spaces, yet there are few mechanisms for capturing and storing this data (with appropriate data sovereignty) over longer periods.
- Providing access to storage and compute resources for specific analytic services, or more generally providing "compute to data" capabilities for sensitive or "too big to download" data.

These activities certainly fall within the scope of a Digital Ecosystem but may still be seen as outside the scope of the Data Space itself or any evolutionary data space initiatives.

A final important dependency would be the variety of services, such as catalogue and marketplace services, or more generally "data intermediation services" as defined by the Data Governance Act, that would need to be provided to enable a Data Space itself to function. Responsibility for operating any such services must be defined and resourced in the planning for the Digital Ecosystem.

2.7. Relationships Between Data Spaces

Building on the DSSC's definition of a data space, any relationships between two data spaces must recognize that each data space will have its own governance framework, which may not be compatible or interoperable with any other data space's governance framework. These gaps in interoperability could appear at every level of interoperability (as defined by the European Interoperability Framework), from technical to organisational, all of which would be "governed" by the chosen governance framework. The Data Sharing Coalition addressed this in [34] and concluded that inter-data space relationships could be achieved in one of two ways:

- The two data spaces jointly took steps to "harmonise" their governance frameworks, for example by agreeing to support or adopt common semantic interoperability frameworks.
- The two data spaces could potentially communicate using proxies or "translators" if automatic conversion of formats or metadata standards were possible.

Allowing the participants of one data space to see, access and potentially use data included in another data space would need to be explicitly authorised by the relevant data providers and data rights holders in the second data space.

- This could be handled through initial agreements by data providers in one data space to allow some or all their data to also be presented in one or more additional data spaces. This approach might be appropriate for open research data, where control over access and use is less important than achieving wide availability and re-use of such data.

Without such an agreement, requests for data in other data spaces may have to be handled through other mechanisms, as long as they respect the data sovereignty of the data providers and data rights holders.

A final contextual consideration would be the possibility of each data space exposing its governance framework in a machine-readable format. This would allow other data spaces to automatically evaluate if the first data space's conditions for data sharing were compatible with their own, opening the door to wider data availability and greater sharing of data among different data spaces.

3. Introduction to Governance

As presented above, the Data Spaces Support Centre (DSSC) defines a "data space" as "*an infrastructure that enables data transactions between different data ecosystem parties based on the governance framework of that data space*". The governance framework of any data space is therefore an essential characteristic of that data space, positively or negatively affecting the range of data that might be made available, what can be done with it, and with whom it can be shared or exchanged. For the GDDS, the needs and requirements of the relevant "data ecosystem parties" – i.e. the community of practice for the European Green Deal – must be identified and addressed through the governance framework proposed for the GDDS.

3.1. Definition of Governance

The term "governance" is defined by the DSSC: "*The creation, development, maintenance and enforcement of a governance framework for a particular scope*". The DSSC defines "governance framework" as "*the set of principles, standards, policies (rules/regulations) and practices that apply to the governance, management, and operations within a particular scope (e.g., a data space, a data space initiative, or data spaces blueprint) as well as to the enforcement thereof, and the resolution of any conflicts*". Unfortunately, these definitions do not address the role of decision-making in governance, such as who should make them and how, or the process for creating the entities that are being governed.

GovLabs notes that "governance is about decision-making"[37] and adapts an OECD definition[38] of governance as "the 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." Both the [32] and [37] recognize that governance and related governance frameworks must operate over multiple layers and possibly through a "governance continuum". [37] highlights that governance "emphasises networks of decision-making across multiple levels."

Considering these and other views of governance (and government) as they can be applied to political entities (nations, states), commercial entities, as well as digital entities such as "digital platforms" and "data spaces", we propose the following definition:

"GOVERNANCE is the process for making decisions about an entity:

[37] Fritzenkötter, J., Hohoff, L., Pierri, P., Verhulst, S.G., Young, A., and Zacharzewski, A., 'Governing the Environment-Related Data Space'. *TheGovLab*, 2022, https://files. thegovlab.org/erdgovernance.pdf.

[38] OECD. (2011). Water governance in OECD countries: A multi-level approach (OECD studies on water). Paris: OECD Publishing. http://dx.doi.org/10.1787/9789264119284-en

- Choosing the questions that must be decided, such as the mission and objectives of the entity, the problems to be solved, and the needs to be addressed.
- Agreeing on the "scope" and "boundaries" of the entity, both initially and over time.
- Ensuring compliance of the entity with applicable laws and regulations.
- Deciding who should be involved in decision-making, including both the actual decision process (including activities like voting, etc.), as well as consultation about each decision.
- Decisions include those about the creation of the entity, such as its form and the relationships between and among outside parties with the entity, as well as who should participate in both decision making and governance and in the operation of the entity.
- Managing the decision-making process, recording both results and details about how these results were decided, such as who was consulted.
- Communicating about the governance process -- identifying who is involved, what decisions are being considered, what decisions have been made.
- Tracking the decisions made, monitoring compliance with these decisions, enforcing those decisions consistent with processes (which have also been decided through the governance process).
- Measuring and reviewing the performance of the entity against agreed objectives."

This definition is consistent with the ideas about governance offered by SITRA [33], Data Sharing Coalition [34], the IDSA Rulebook v2.0 [39], and OpenDEI's "Design Principles for Data Spaces"[40]. "Design Principles for Data Spaces" [41] introduces the concept of "soft infrastructure": "the suite of agreements that enable a data space to function – this is the governance and glue that enable a system of systems to successfully operate".

3.2. Translating Governance Requirements into Implementation – Dimensions of Implementation

Based on the definition above, governance involves the translation of the objectives of, the problems to be solved by and/or the needs to be addressed by an entity into decisions about how to meet those objectives, whether legally, organizationally, and/or technologically[41]. The European Strategy for Data (referenced at [5]) describes data spaces as data ecosystems with shared legal, operation, functional agreements and technical standards, so such categories for implementation are appropriate.

To illustrate, for a government, governance involves ongoing identification of problems or needs, decisions about how to address those needs, implementation of the agreed response through various means (regulation, taxation, financial and human support programs, legislation), and measurement of its effectiveness (sometimes through election of different leaders). Governments themselves are often formed as part of a governance process, with decisions about the form of government, the different "organs" of governance, and their relationships, decided through a collective process and documented in a "constitution".

39 IDSA Rulebook v2 https://docs.internationaldataspaces.org/ids-knowledgebase/v/idsa-rulebook/front-matter/readme
40 Nagel, L. and Lycklama, D. (Eds), 'Design Principles for Data Spaces'. International Data Spaces Association, 2021, doi:10.5281/zenodo.5244997
41 EUHubs4Data: Evaluation and recommendations on the legal conditions for trading data in a complex ecosystem https://cordis.europa.eu/project/id/951771/results

For a corporation, governance ensures the organisation is working toward agreed objectives (including profit for a for-profit corporation), using the agreed strategy, while complying with both internal policies and external rules. Initial governance decisions relate to establishing the rights of shareholders (or members), voting procedures, the respective roles of the board/executive council vs. members and vs. executive management, etc., all of which are documented in the corporation's charter or articles of incorporation and accompanying by-laws. Most for-profit corporations delegate operational and implementation questions to management in order preserve flexibility, focussing governance activities on supervision of management, reviews of strategy and performance and resolution of stakeholder disputes.

For a data space, governance involves the initial identification of stakeholder problems and needs, decisions about how to meet those needs (e.g., with legal agreements, operational processes, and/or technology), followed by implementation and performance monitoring. For a technically oriented undertaking like a data space, there are strong parallels between "governance processes" and "requirements analysis".

Governance requirements identified in Chapters 4 (generically) and 5 (for the Green Deal specifically) will be highlighted and numbered to allow them to be tracked through various mechanisms, not only in the proposed GDDS Governance Framework, but also in its Technical Blueprint and Implementation Roadmap.

3.3.Stages of Governance – The Lifecycle Dimension of Governance

Governance has a lifecycle, as highlighted by our proposed definition of governance, as well as by several observers[42] [43]. Governance starts with agreement among a group of initial stakeholders ("founders") that an "entity" should be created to address a common need and progressively translating that initial agreement into the concrete formation of the entity and its associated governance. There are certain governance steps to be taken during the initial formation of the entity, including making decisions about how the entity should be designed and operated. After formation, the entity begins to operate, requiring operational activities, and is expected to fulfil its mission, requiring monitoring and enforcement activities, as well as continuous improvement and innovation. To be complete, the governance lifecycle should consider the steps needed to ensure the entity's sustainability and persistence, or possible merger, combination or termination.

Stages of governance therefore include:

- Formation
- Operation
- Monitoring and Enforcement, Continuous Improvement
- Sustainability.

In many cases, decisions taken at the Formation stage drive corresponding requirements for operations as well as monitoring in support of both enforcement and continuous improvement.

[42] Lis, Dominik & Otto, Boris. (2021). Towards a Taxonomy of Ecosystem Data Governance. 10.24251/HICSS.2021.733.
[43] ISO/IEC 38500:2015 https://www.iso.org/standard/62816.html

We itemise these operations and monitoring requirements in connection with a few Formation requirements, but in many cases, these are clear and are not presented explicitly.

In some cases, we identify "KEY DECISIONS" that must be taken for the data space. For example (see section 5.1.1), if any data in a data space might have limits on access and use, the data space will need uniform rules and procedures for working with such data, which would need to be put in place at the beginning of the data space's life and would impact not only governance but also the functional, legal and technical requirements for the data space.

Similarly, we identify the need for definitive "GUIDANCE" on various subjects, for example on the practical interpretation of certain legislation. This guidance could be provided by the DSSC, or later from the European Data Innovation Board (EDIB). Such guidance would be important for developing a coherent governance framework.

3.4.Layers of Governance – the Context Dimension of Governance

The governance framework for a data space has multiple layers. This layered approach has been proposed by several analyses, including [40] and Torre-Bastida, et al[44].

We analyse governance requirements in four layers, as follows:

- **Legal Compliance Layer:** This layer translates relevant legislation and regulation into governance requirements that apply to one or more of the other three layers: digital platform governance, data space governance or data governance.
- **Digital Ecosystem and Platform Governance Layer:** This layer addresses the good governance of a community of practice and related digital infrastructures supporting specific objectives. Digital platforms come in many forms, and many connect providers and users of resources (such as data or services) through a so-called "two sided market".
- **Data Space Governance Layer:** In this layer the generic capabilities of a digital platform are focused on enabling and supporting the "data transactions" contemplated by the DSSC in its definition of a data space, as well as the value added services contemplated by the data space definitions offered by the DGA and DA. Notably the Data Space Governance Layer captures the processes required for data sovereignty, translating information provided by data holders and data rights holders, including the subjects of personal data, in the data governance layer into safe and approved access and use of that data.
- **Data Governance Layer:** This layer ensures transparency in the presentation and potential use and access for each item of data, enabling not only data sovereignty, but also compliance with a growing range of relevant legislation.

Some analyses of governance present the governance structure that might be agreed by participants in broad initiatives such as Gaia-X or the International Data Spaces Association as a separate layer. We attempt to incorporate the insights into governance provided by such initiatives but do not segregate them.

[44] Torre-Bastida, A.I., Gil, G., Miñón, R., Díaz-de-Arcaya, J. (2022). Technological Perspective of Data Governance in Data Space Ecosystems. In: Curry, E., Scerri, S., Tuikka, T. (eds) Data Spaces. Springer, Cham. https://doi.org/10.1007/978-3-030-98636-04

3.4.1. Legal Compliance Framework.

The first layer is the general legal and regulatory framework for any "data space initiative". For data spaces based in the EU, EU and Member State legislation and regulation determines this framework. Figure 3 (from the DSSC) identifies the range of horizontal EU legislation that applies to data spaces and data space initiatives.

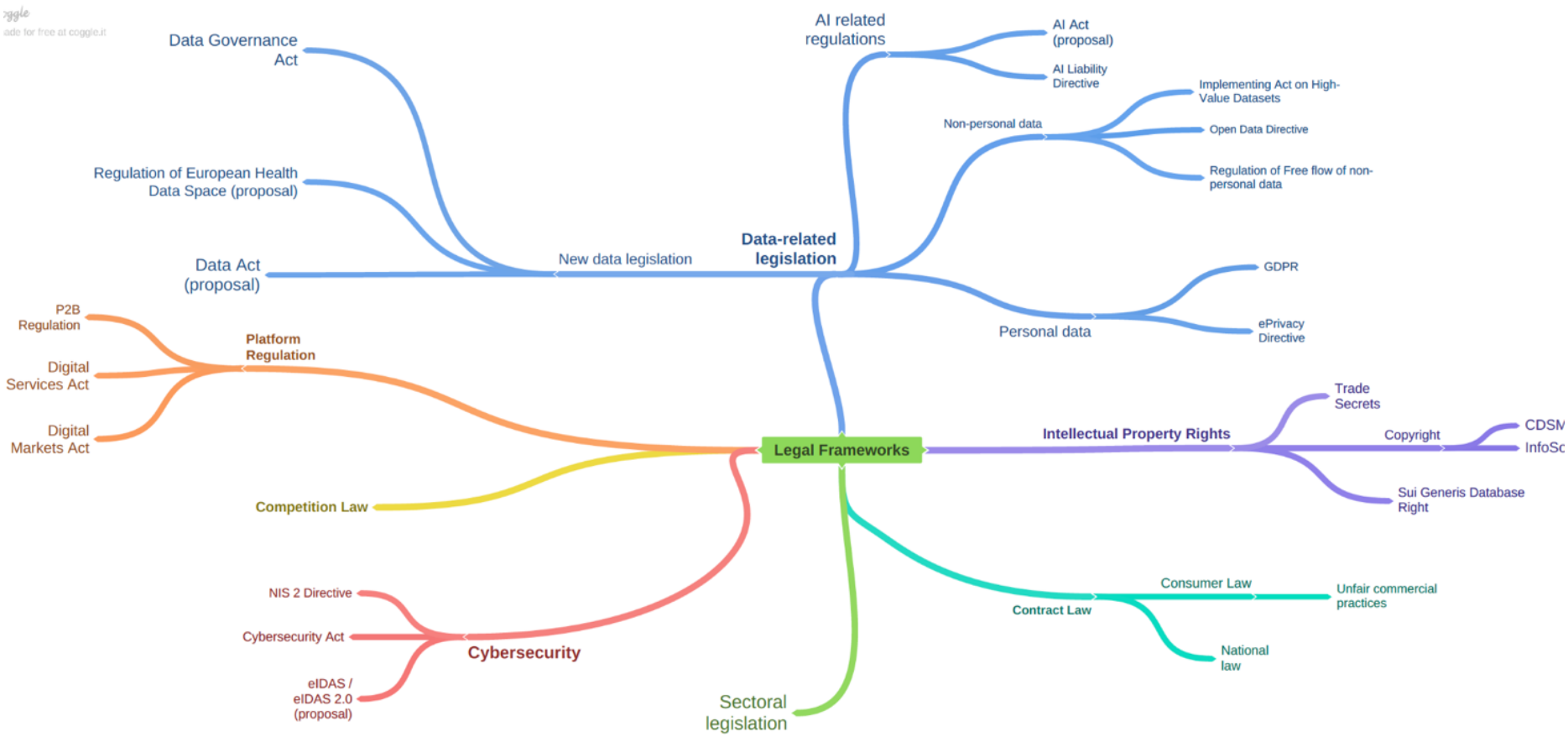

Figure 3: Horizontal EU Legislation Relevant to Data Spaces[45]

[45] Figure provided by the DSSC Legal Expert Team

It includes EU requirements for any "digital platform" initiative (such as the Digital Markets Act and Digital Services Act), as well as in force and proposed legislation targeting "data sharing" and "data transactions," including the General Data Protection Regulation (GDPR), covering personal data, the Data Act, covering non-personal data, possible regulatory roles such as the European Data Innovation Board (EDIB) and several other instruments and roles. Legal requirements mapped to this layer translate into specific requirements that must be implemented in other layers to be compliant with relevant legislation. For example, the Data Governance Act requires that certain "data intermediation services" must be provided by distinct legal entities, creating a requirement mapped to the Data Space Governance layer, where data sharing is addressed.

3.4.2. Digital Ecosystem and Platform Governance

The next layer is a general framework for governance of multi-stakeholder digital ecosystems and related digital platforms. This layer organises the community of stakeholders that are working together to reach a range of shared objectives. While it is referred to as a "governance layer" it does not directly bind participants to specific rules, but it enables organisations to work together through shared principles and use of best practices. There may be multiple digital platforms or infrastructures created within this ecosystem, and each one should reflect these common principles and practices, thereby enabling interoperability between them. Our Generic Governance Framework thus builds on the idea that a data space would be implemented as a digital infrastructure, or "digital platform"[46].

Recalling our definition of "governance" above, which is concerned with decisions about a specific "entity", there may be several types of related "entities" of concern in this layer:

- The organising legal entity that takes on the "personality" of a specific digital platform, leading the governance process with stakeholders, and entering into contractual agreements on behalf of the digital platform (for example with participants or key service providers).
- The digital platform itself, including the roles, functions, procedures, and technology that define the platform and its operations.
- One or more legal entities, acting as "data intermediaries", delivering "data intermediation services" as defined by the Data Governance Act (see Appendix I for further discussion).
- One or more legal entities, such as Data Altruism Organisations and Data Cooperatives (as defined by the Data Governance Act), acting on behalf of collections of data providers and data rights holders to provide their data collectively for either altruistic (public good) purposes or for the collective benefit of the participating data providers.

These various entities may be related, but the differences between them can be significant. For example, the organising legal entity may define "members" and "membership" so that members

[46] While the DSSC defines a data space as "an infrastructure that enables data exchange transactions…", the DSSC does not define "infrastructure" and does not specify that the infrastructure required for a data space must be "digital". In fact, a traditional library, with printed books, shelves organised by category, and a card catalogue by subject, is every bit as much a "data sharing infrastructure" as any digital infrastructure. "Data exchange transactions" have been executed and recorded with writing or printing on paper for centuries. Thinking about how "analogue infrastructures" can be used to implement data exchange transactions can offer insights into the principles required for governance at different levels.

can engage in decision-making about the entity operating the platform. There will also be "participants" in the digital platform, typically providing and/or consuming data or services, as well as participants that perform services essential to the operation of the platform. These two roles are distinct: participants in the platform may or may not be members of its organising entity, and members may or may not act as participants. It is likely that most participants will want to have a "voice" in governance, to express their views on the decisions being made, but it is not practical to give them a direct "vote" on every matter.

In addition to the different objectives and functions performed by the different entities, each legal entity would need to select a suitable legal form, either established in a specific Member State (such as an AISBL established in Belgium) or using one of a number of forms that can be established on a pan-European basis (such as the European Digital Innovation Consortium – EDIC).

Best practices for digital platform governance have been identified through analysis of literature on the governance of such platforms. These have been translated into specific requirements in section 4.2, including defining the principles and values that the digital platform and its organising entity will be designed to follow and respect.

Section 4.1.2 highlights the need for guidance on best practices for identifying and forming appropriate legal entities. Section 4.2.1.5 explores requirements for creating responsive governance mechanisms, that give stakeholders a voice in governance, while ensuring effective decision-making. These requirements reflect the best practices in multi-stakeholder governance presented in Appendix II.

3.4.3. Data Space Governance

In the context of the digital platform and the entity governing that platform, specific decisions must be made about how to enable the data transactions that the infrastructure will support. Decisions include specifying:

- Data Space-specific roles (e.g., data providers, data consumers).
- Data Space-specific rules for each role.
- Specific services the platform must perform with a certain level of quality to enable data transactions (e.g., catalogues, controlled vocabularies and associated services, authentication/authorisation, data transport, …).
- Mechanisms for tracking adherence to agreed principles or legal requirements, such as fairness, reasonableness, non-discrimination, transparency, etc.

It may be possible to design and implement a specific digital platform to support multiple data spaces. This is a common architectural approach for many software-as-a-service capabilities. Ignoring questions of "where" such a system might be instantiated, related data sovereignty concerns about where the collective metadata of a data space would be stored, and where the technical processes of the data space would be executed, it will be important for such a technical system to be configured to accommodate the governance requirements specified at all levels of the governance framework, and to allow the different tenant data spaces to be configured as specifically required by their specific governance frameworks and corresponding data ecosystems.

Similarly, a single data space might be implemented with a distributed technical architecture where independent systems are compliant with common specifications and interact using agreed

interfaces and protocols. In addition to assuring compliance and technical interoperability, these independent systems must also be configurable to implement the required governance framework of the data space. When a data space is being implemented, its founding members will make the governance decisions required to achieve the objectives of their community, and technical choices should be made that will accommodate those governance decisions, rather than the other way around.

3.4.4. Data Governance.

In this layer, "data governance" – that is, decisions about the visibility, access, and usage of specific pieces or sets of data – must be addressed as part of the overall governance framework. Just as there may be stakeholders concerned with the governance of the digital platform, and who may have a role in that governance, there are also stakeholders for the governance of each item of data, such as data subjects, data holders, holders of intellectual and other rights, etc., who should also have a role in decisions about whether and how each item of data can be seen, accessed, used, and re-used.

In general, each item of data is made available to one or more data spaces according to the access and use policies set by the data holder and/or data rights holders. However, that data holder will need assurance from each of those data spaces that they will respect and implement the access and use policies required by the holder. Before data spaces are created, decisions should be made about the planned regime for data access and use policies, so that the data space can accommodate the needs of the data holders that will be asked to participate in the data space. Once a data space has been created, data holders will assess the data space's chosen data governance scheme to decide whether to provide data to that data space.

4. Generic Governance Framework

This chapter presents a generic governance framework that would be suitable for any pan-European data space, regardless of sector or domain. It is expressed as a concrete set of governance decisions and requirements, flowing from considerations highlighted in the various treatments of data space governance mentioned above. Taking inspiration from Sitra's "Rulebook for a Fair Data Economy", we have attempted to synthesise those qualitative considerations into a complete list of governance decisions and requirements that must be addressed by most, if not all, data spaces.

A comprehensive framework such as this could be used to describe governance decisions made for each individual data space and could allow different governance frameworks to be compared to assess compatibility and interoperability. For example, even without "joining" a particular data space, a data holder considering providing data to that data space could assess whether its governance framework is appropriate for the kind of data that might be provided. If the data holder holds data about critical infrastructure, and a data space's governance framework declares that "no data that identifies individuals or describes critical infrastructure will be included or processed in this data space", the data holder knows that its data should not be provided to that data space.

4.1.Legal and Regulatory Context

This section organises requirements for governance flowing from two categories of legal and regulatory instruments: horizontal across the EU, and Member State specific. Legal requirements that are specific to the Green Deal are addressed in the next chapter.

4.1.1. EU Horizontal Legal and Regulatory Context

Appendix I presents a review of several "horizontal" EU laws and regulation (see Figure 5 above for an illustration) of relevance to data space governance. Here we summarise how the analysed legislation translates into requirements for a Generic Governance Framework. Compliance defines several requirements that map to various layers of governance: "data governance," "data space governance" and "digital platform governance".

4.1.1.1. EU Requirements for Data Governance

At the level of Data Governance, transparent declaration of requirements for processing, including access and use, are required to ensure compliance with relevant legislation, as follows:

Tracking Data Categories, Data Lifecycle Stages and Related Compliance Requirements: Prior to the latest suite of new legislation, the primary instrument affecting data governance was the GDPR. Some infrastructures[47] have used the W3C's Data Privacy Vocabulary[48] to track personal data inside their infrastructure. This consistent information enabled the infrastructure to comply with the GDPR. For example, the DPV captures the purpose for which the personal data was collected, the legal basis for that collection, and whatever mechanisms might be available for requesting consent from the data subject for any other uses proposed for the data. This

[47] Hernandez, J., McKenna, L., Brennan, R. (2022). TIKD: A Trusted Integrated Knowledge Dataspace for Sensitive Data Sharing and Collaboration. In: Curry, E., Scerri, S., Tuikka, T. (eds) Data Spaces . Springer, Cham. https://doi.org/10.1007/978-3-030-98636-0_13

[48] https://w3c.github.io/dpv/dpv

information in turn determined what could be done with the data, and which processes would be executed.

New legislation flowing from the EU Strategy for Data, such as the Data Governance Act and Data Act, as well as older legislation such as the NIS Directive, identify similar, but different, compliance requirements related to a number of other categories of data. This includes data held by public sector bodies that may not be made public because it contains confidential information, Internet-of-Things data, data used to train AI models, data about critical infrastructure, etc. This legislation also defines different compliance requirements at different stages in the lifecycle of data, including storage and processing, visibility, findability, accessibility, interoperability, and re-use.

If the data sharing infrastructure will exclusively refer to and work with data that is 100% open, without any restrictions on re-use, the identified requirements do not apply. However, if the infrastructure might include data with any limits on access or re-use, the infrastructure will need consistent mechanisms for tracking the status of data of each category across its lifecycle. These mechanisms could affect all aspects of infrastructure implementation, from business model to legal framework, to operations, to functional requirements and technology.

GGF-1.01/KEY DECISION: If the infrastructure will work with any data with any limits on access or re-use, the following requirements must be met for compliance with the Horizontal EU Legal Framework.

GGF-1.01/GUIDANCE: Create a "Data Protection" taxonomy and vocabulary describing the multiple categories of data defined by horizontal data legislation, potentially including both EU and individual Member State requirements, as well as the multiple lifecycle stages of that data, that may be supported by a data space, capturing the required data governance model(s) for each category and lifecycle. This could build on or extend the W3C: Data Privacy Vocabulary[49] which currently addresses only personal data and the GDPR. Legal review of the Taxonomy and related process requirements would be needed to test completeness and adequacy.

GGF-1.01/DG/Formation: Incorporate an extended Data Protection Taxonomy into the design of the data space infrastructure, requiring inclusion of this information in all metadata in data intermediation services such as catalogues, as well as support for related processes (e.g., requesting consents from data subjects) that would be triggered by this metadata. Roles, requirements, functions, rights, obligations, and performance expectations, along with associated onboarding and performance monitoring processes, are defined.

GGF-1.01/DG/Operation: Establish and operate the necessary processes for working with each category of data, at each stage of its lifecycle. Proper application of the Data Protection Taxonomy to each item of data, combined with availability and use of corresponding processes and workflows, including tracking of these processes, should yield compliance with the relevant legislation and regulation encoded in the Taxonomy.

GGF-1.01/DG/Monitoring-Compliance (M-C): Compliance with requirements by category is tracked, reported and aggregated. Audits of compliance are conducted, and non-compliance[50]

[49] https://w3c.github.io/dpv/dpv/

[50] For example, failure of a data holder to correctly label data as falling into a specific data category, allowing use of the data without following the correct procedures.

is addressed through agreed procedures, including the possibility of data being removed, suspended from the data space, or marked as non-compliant.

GGF-1.01/DG/Monitoring-Improvement (M-I): Opportunities for improvement are identified and implemented. Role definitions, requirements, functions, rights, obligations, and performance expectations, along with associated onboarding and performance monitoring processes, are reviewed periodically and revised as needed for the data space to meet its objectives.

Collective Data Providers: Data Altruism, Data Cooperatives, etc.: The Data Governance Act defines both Data Altruism Organisations ("DAltOs") and Data Cooperatives. As defined, DAltOs are not treated as Data Intermediaries, while Data Cooperatives represent a form of Data Cooperative. DAltOs represent an EU-recognized form of data altruism organisation that might also be created under the jurisdictions of various Member States (MS). DAltOs, MS-recognized data altruism organisations and Data Cooperatives all share attributes with a range of data governance models that have been identified in the literature, including "data sharing pools", "data cooperatives", "public data trusts" and "personal data sovereignty"[51]. These various kinds of entities might all take on the roles of data provider, data holder and data rights holder, acting on behalf of groups of data subjects and/or non-personal data holders. It is unclear whether the forms defined by the Data Governance Act offer benefits over other forms of collective data holder. In addition to direct responsibility for acting on requests for access and use of data in their custody, these organisations might also act as stakeholders in governance, representing, for example, the interests of citizens generating data of relevance to the Green Deal and other broad data sharing initiatives of interest to the general public.

GGF-1.02/GUIDANCE: Develop best practices for the use of Data Altruism Organisations, Data Cooperatives, and other forms of collective data provider, including the pros and cons of each form of entity. Consider their effectiveness as representatives of the ultimate data subjects and non-personal data holders, the legal, operational and technical measures required for them to act on behalf of those contributing data, and their ability and suitability to represent the interests of their contributors in governance decision-making.
GGF-1.02/KEY DECISION: Based on the Guidance, decide whether one or more collective data providers should be established to assist in the execution in various use cases targeted by the data space in question.

4.1.1.2. EU Legal Requirements for Data Space Governance

At the level of Data Space Governance, compliance requires the following actions:

Tracking Data Service Types and Related Compliance Requirements: The Data Governance Act distinguishes three kinds of services: data intermediation services performed by Data Intermediaries, other services that may be performed by Data Intermediaries, and services that

[51] E.g., Micheli, M., Ponti, M., Craglia, M., & Berti Suman, A. (2020). Emerging models of data governance in the age of datafication. Big Data & Society, 7(2). https://doi.org/10.1177/2053951720948087

should not be bundled by Data Intermediaries with Data Intermediation Services. The Artificial Intelligence (AI) Act (proposed)[52] will also place requirements on services for training and using advanced analytics capabilities. To manage compliance of these services with the relevant legislation, a taxonomy of different service types should be created and rules extracted from the legislation to guide practical operation of the data space.

Note that the proposed AI Act places additional requirements on AI systems, which "*means a machine-based system that is designed to operate with varying levels of autonomy and that can, for explicit or implicit objectives, generate outputs such as predictions, recommendations, or decisions, that influence physical or virtual environments*". Various value-added services, such as data analytics and data fusion systems might easily be categorised as AI systems with this definition, and some might be categorised as "high risk" AI systems, requiring additional requirements and protective measures. These requirements should be captured in this action.

GGF-1.03/GUIDANCE: Create a Data Service Vocabulary (and taxonomy) to label services covered by the Data Governance Act, AI Act, and any other acts defining regulated data services. Define the required service governance model(s) for each type of service (e.g. how a provider can show that it does not bundle data intermediation services with other prohibited services) and the rules that should be applied to ensure compliance (e.g. data analytics services for items of data may not be provided by the Data Intermediary that arranges a data transaction for those items of data). Legal review of the Service Taxonomy and related process requirements to test completeness and adequacy.
GGF-1.03/DSG/F: Incorporate the Data Service Vocabulary into the design of the infrastructure, requiring inclusion of this information in all metadata in service catalogues, as well as support for related processes (e.g., tests for fairness, reasonableness, non-discrimination, as well as determination of special status of users, e.g. small or micro enterprises). Roles, requirements, functions, rights, obligations, and performance expectations, along with associated onboarding and performance monitoring processes, are defined.

Data Intermediaries: The Data Governance Act presents a number of requirements for Data Intermediaries and Data Intermediation Services (see Appendix I). It is difficult to translate the DGA's requirements into clear recommendations on those circumstances where Data Intermediary treatment is appropriate and those that might avoid this treatment. There is also confusion between the DGA's apparent limitations on offers of value added services, much less the integration of such services, and the interpretation of the DGA by others, such as the JRC[53] and [32], that data intermediaries should play a role in generating high value datasets, integrating data from multiple sources, and creating value-added data products that nevertheless respect data sovereignty.

GGF-1.04/GUIDANCE: Develop best practices for applicability, implementation and operation of data intermediaries as defined by DGA. (Appendix I identifies a range of questions where advice and interpretation of the DGA are needed.)

[52] https://www.consilium.europa.eu/en/press/press-releases/2022/12/06/artificial-intelligence-act-council-calls-for-promoting-safe-ai-that-respects-fundamental-rights
[53] Kotsev, A., Escriu Paradell, J. and Minghini, M., Beyond INSPIRE. Perspectives on the legal foundation of the European Green Deal Data Space, European Commission, 2023. https://publications.jrc.ec.europa.eu/repository/handle/JRC133958

GGF-1.04/KEY DECISION: Based on the Guidance, decide whether Data Intermediation Services are being provided as regulated by the Data Governance Act and how to organise a separate legal entity (or multiple entities) to provide such services.
GGF-1.04/DSG/F: Include/respect the legal definition of Data Intermediary and Data Intermediation Service(s) in the design of the infrastructure. Establish an appropriate legal entity as a Data Intermediary to operate the Data Intermediation Service(s). Register with the competent legal authority in the appropriate jurisdiction.
GGF-1.04/DSG/Operation: Operate the Data Intermediation Service(s).
GGF-1.04/DSG/Monitor-Compliance: Collect data on the operation of the Data Intermediation Services as required to fulfil requirements specified for compliance with the DGA (e.g., proving that services were provided in a fair, reasonable and non-discriminatory manner).
GGF-1.04/DSG/Sustainability: Plans are established and maintained to ensure continuity for any Data Intermediation Services.

Unfair Terms For Business–To-Business Data Sharing. Article 13 of the recently approved Data Act prohibits the "unilateral" imposition of "unfair terms" for data sharing. While *prima facie* reasonable, this Article may limit the ability of data space initiatives to establish standard terms and conditions for participants, for example, through a standard Constitutive Agreement.

GGF-1.05/GUIDANCE: Can standard terms and conditions be used to govern participation in a data space initiative, e.g. through Constitutive and Accession Agreements (as proposed by Sitra[54])?

Interoperability in Data Spaces and Data Sharing Contracts. Article 28 of the recently approved Data Act establishes a number of requirements for data providers regarding transparent annotation of their data to facilitate finding, accessing and using that data, declaration of related interoperability standards (such as data structures, vocabularies, etc.), automated data access mechanisms and interoperable data sharing contracts.

GGF-1.06/DSG+DG/F: Include the requirements of Article 28 in the specific governance requirements in this framework:

- *GGF-3.1.01/DSG/F* Domain Data Models and Interoperability Standards
- *GGF-3.1.06/DSG/F*: Creating a Framework for Declaring Access, Usage Policies
- *GGF-4.1.02/DG/F*: Define Metadata/Self-Descriptions required for all data
- *GGF-4.1.03/DG/F*: Define Measures of Quality and Fitness for Purpose

4.1.1.3. EU Legal Requirements for Digital Platform Governance

At the level of Digital Platform Governance, compliance requires the following action:

Cybersecurity Required for Personal Data and/or Data About Critical Infrastructure. If data sharing infrastructure might refer to and work with data that is either "personal", as defined by the GDPR, or related to "critical infrastructure", as defined by the NIS Directive, the data sharing infrastructure must be designed and operated to ensure the legally required levels of cybersecurity

[54] See [33] Constitutive and Accession Agreements

for the data. These mechanisms could affect all aspects of infrastructure implementation, from business model to legal framework, to operations, to functional requirements and technology.

GGF-1.07/KEY DECISION: If the infrastructure will work with any data that legally requires the use of secure digital infrastructure, the following requirements must be met for compliance with the Horizontal EU Legal Framework.
GGF-1.07/GUIDANCE: Provide best practice recommendations for creating digital infrastructure that is suitably secure to hold, store, process and otherwise manipulate data that requires such infrastructure.
GGF-1.07/DP/Formation: Define processes, roles, requirements, functions etc., including technology requirements, required to establish a secure infrastructure for the storage and processing of sensitive data. Implement the technology platform in compliance with the defined requirements.

Appendix III explores best practices for operating secure IT facilities that cover additional requirements for operations, compliance and performance improvement.

4.1.1.4. Other EU Legal Requirements for Governance

Our analysis has not considered the full range of horizontal legislation identified in Figure 5, including:

- Digital Markets Act
- Digital Services Act
- AI Act (proposed)
- Platform-to-Business Regulation (P2B)
- Geo-blocking Regulation
- Copyright Directive
- Database Directive
- Antitrust and Competition Laws
- EU Consumer Protection Directive
- Payment Services Directive 2 (PSD2)
- e-Commerce Directive.

Each of these laws may contain requirements that should be captured in this Generic Governance Framework.

GGF-1.08/GUIDANCE: Develop best practices for common European Data spaces to properly comply with the range of horizontal legal and regulatory requirements that apply throughout the EU. Translate these practices into requirements at the different levels of governance, which can be implemented through legal, operational, and/or technical mechanisms.

4.1.2. "Choice of Law" Clauses and Member State Legislation and Regulation

As discussed on Appendix I, whether explicit or not, "data transactions" enabled by a data space are covered by the laws and regulations of one or more Member States. This could be through the "choice of law" clause of specific data exchange agreements between two parties, through the law selected to govern the data space in which such a data transaction occurs, through legal claims

of one or more affected parties, and possibly through enforcement efforts by one or more Member States. These different mechanisms of jurisdiction are possible simultaneously – providing a "choice of law" clause in a specific data exchange agreement does not prevent other mechanisms being applied or other jurisdictions becoming involved.

Each Member State decides how EU Directives are implemented ("transposed") at the national level, and those laws may be supplemented or modified by national laws and regulations. Even EU Regulations, which establish consistent rules across all Member States, can specify "derogations" to Member States to accommodate their existing legislation or practices. For example, the GDPR applies to all organisations that process personal data about EU citizens, regardless of where the organisation is based. However, the GDPR also allows for some flexibility in how it is implemented, which has led to variations in how it is applied between Member States. The GDPR allows Member States to introduce their own specific rules on data processing in certain areas, such as health, employment, and research, which can lead to differences in how the GDPR is applied in these sectors. In the health sector in particular, personal data must be handled very differently, depending not only on the applicable jurisdictions, but also the specific type of health data in question (e.g., genetic, biometric, etc.).[55]

GGF-1.09/GUIDANCE: Develop best practices for common European Data spaces to properly comply with the range of legal and regulatory requirements that apply in different jurisdictions across the EU. Provide recommendations regarding suitable "choice of law" jurisdictions to be specified in the various legal agreements that will be required to establish common European Data Spaces, with participants from multiple jurisdictions and supporting cross-border data transactions.
GGF-1.10/GUIDANCE: Where one or more legal entities are required, provide recommendations regarding suitable forms of entity (e.g. an EDIC, AISBL in Belgium, European Association, etc.), and optimum jurisdictions in which to establish such entities.

4.1.3. Sector-Specific Legislation and Regulation

Legislation applicable at the sectoral level is identified at the use case level depending on the data, stakeholders and sector involved. Requirements are highlighted here for the INSPIRE and the High Value Datasets regulations discussed in Appendix I. These apply primarily at the Data Space Governance layer since they address related services and data models/interoperability standards, as well as the Data Governance layer requiring the use of standardised metadata. While they are very relevant to the Green Deal Digital Ecosystem, they apply to other common European Data Spaces as well.

GGF-1.11/DSG+DG/F: Public administrations should apply the INSPIRE Implementing Rules (IR)[56] for geospatial datasets and should make available High Valuable Datasets free of charge in machine readable format via APIs and, where relevant, as bulk download. Other data and service providers should consider aligning with the corresponding requirements. The IRs map to the following specific governance requirements in this document: ● *GGF-3.1.01/DSG/F* Domain Data Models and Interoperability Standards ● *GGF-3.1.02/DSG/F*: Intermediation/ Marketplace/ Catalogues

[55] Kiseleva, A. and De Hert, P., 'Creating a European Health Data Space. Obstacles in Four Key Legal Areas'. *European Pharmaceutical Law Review*, Vol. 5, No 1, 2021, pp. 21–36, doi:10.2139/ ssrn.3846781.

[56] https://inspire.ec.europa.eu/inspire-implementing-rules/51763

- *GGF-3.1.03/DSG/F*: Ancillary Services: Data Preparation, Encryption, Anonymization, Transformation
- *GGF-3.1.04/DSG/F*: Enrichment, Aggregation, Fusion, Analysis, AI/ML
- *GGF-4.1.02/DG/F:* Define Metadata/Self-Descriptions required for all data

4.2.Generic Digital Platform Governance

Many important aspects of data space governance identified by the data space community of practice have previously been identified in broader analyses of the governance of digital platforms, which are not limited to enabling data transactions, and can be operated either within an organisation through its information technology (IT) efforts or on behalf of multiple stakeholders.

One feature of digital platforms is their organisational nature (or "configuration", per the taxonomy of Ecosystem Data Governance, referenced at [42]), falling into one of several categories:

A. Dominated by a single player (e.g., Google or Apple), which then enables different forms of participation in the platform (both organizationally and technologically). Organisational participation is typically structured around contracts between participants and the dominant player.
B. Formal multi-organization alliance, where the alliance is represented by a central entity, often a legal entity, with a defined governance framework and formal mechanisms to establish and govern the technical architecture adopted by platform. That technical architecture could be centralised, distributed, or a hybrid of the two.
C. Distributed alliance of multiple organisations or participants. Both organisational and technological structures are distributed and decentralised. Most platforms implementing blockchain and/or similar technologies fall into this category.

Our analysis focuses on category B, a multi-organizational alliance organised around a central entity (the "organising entity")[57]. The organising entity, at minimum, performs organisational governance functions on a centralised basis. Decision-making authority and operational resources can range from centralised to distributed, and technological functions can also be performed as a mix of centralised and fully distributed services.

Different bodies of information explore possible governance requirements from different perspectives, and are summarised in corresponding Appendices:

- Appendix II summarises best practices for governance of multi-organizational alliances (whether they operate a digital platform or not) as well as social enterprises (which create public goods and/or serve the public or general interest). These best practices are integrated as appropriate into the design and governance framework below.
- Appendix III summarises best practices for governance of IT operations, as well as cybersecurity. These best practices are integrated as appropriate into the design and governance framework below.

All these perspectives apply to data spaces, so requirements identified from any of these perspectives will be relevant for data spaces.

[57] Category A, platforms dominated by a single player, includes several examples of "data ecosystems" to which the European Strategy for Data was designed to create alternatives. Category C, fully distributed platforms, is a potentially interesting model, but out of scope for this analysis.

The sections below follow the design/governance lifecycle to present the key dimensions of digital platform governance that must be considered in the creation of a new digital platform.

4.2.1. Generic Digital Platform: Formation

During the Formation stage, governance and design activities work hand-in-hand. As noted in section 2.3, the formation of a new digital platform typically involves the work of a group of interested stakeholders, working on a volunteer basis, to create the digital platform to meet common goals and objectives. The commonality of these goals is the glue that brings this group together.

While this initial process may be iterative, and new stakeholders may be asked to join the working group over time, the group must tackle five key sets of questions for the initial design of the platform:

1. **Landscape Design.** What is the initial structure of the platform, its organising entity, any related entities, and its relationship to other entities and initiatives? Which stakeholders need to have a vote on which issues, and if not a vote, how are their voices heard in decision-making?
2. **Mission.** What is the mission and objective of the platform? What use cases must it enable? What are the objectives of each use case, and how can the platform be set up to enable those use cases to be successful? Who needs to be involved, performing what roles? What resources are needed? What are the rules to be respected within the platform?
3. **Value Creation.** How is value created in the platform, and for whom? This should be examined for each use case, for various participants, and then for the digital platform as a whole.
4. **Technical Architecture.** What is the architecture of the platform? Is it centralised, federated, distributed? How does the architecture map to governance, to responsibility and accountability?
5. **Governance Architecture**. What is the shape of platform governance? What entities must be established, what governance and advisory bodies are needed, what agreements or operating procedures must be created?

Requirements in this section focus on the Formation stage of governance, but related operational, monitoring and sustainability requirements are called out in selected cases.

4.2.1.1. Generic Digital Platform: Landscape Design

While most discussions of data spaces assume independent creation, and a nearly immaculate conception, in reality data spaces are created by participants in an existing or growing community of practice, who want to create an infrastructure to support data transactions that will in turn enable use cases with various objectives and impacts. This creation process must reflect this landscape.

Sections 1.2 - 1.5 in this chapter explore some of these contextual issues. Additional “conceptual” and “relational” design dimensions are listed below, addressing the relations and overlaps between

and among use cases, stakeholders, required data and services and even funding and accountability:

- Do the different use cases that will be considered in the design address similar sets of stakeholders? To what extent is there overlap? To what extent is there clustering or stratification in the landscape?
- Is there overlap among the resources (data, services) needed by the different use cases? To what extent does each use case need unique resources for success?
- For a given use case there are usually key organisational stakeholders – those that hold critical data, those for which the results of the use case are critical, those that hold accountability for those results (e.g. governments, funding bodies). Is there overlap among these key stakeholders, or is there clustering or stratification?
- How will existing, trusted communities of practice participate in the data space initiatives? Are the limits to their ability to operate in a trusted fashion within a larger community? Are there technical solutions to this problem, or do selected use cases and communities need different governance structures to accommodate their needs?
- Are there overlaps with related initiatives, e.g., other data spaces, that require compatible designs and structures?

These questions motivate the following broad question of design and governance:

GGF-2.1.01/DP/F: Landscape Design: How will the infrastructure relate to its community of practice, its "digital ecosystem"? Will multiple infrastructures be needed to accommodate variations in objectives, resources, stakeholders, governance constraints, relationships with external initiatives, etc.? Are there certain design choices that can be made similarly for multiple infrastructures, while other design choices need to be different to accommodate the needs of specific use cases?

4.2.1.2. Generic Digital Platform: Mission and Objectives

Enabling the data value chain lies at the core of the data space digital platform mission and objectives. Table 1 in chapter 1 presents a taxonomy of objectives, organised into a "ladder" of ascending levels as functionality is added to the platform to achieve more ambitious goals.

GGF-2.2.01/DP/F: Objectives: What is the "purpose" of the infrastructure, the problem it is designed to solve, or the objective(s) it is designed to achieve? Describe the "raison d'être" for the stakeholders involved. Examples: supply chain, maintenance services, research and innovation activity, logistics, business cooperative, data marketplace or platform, testbed. [58]

- Best Practice: Align with "Objectives Taxonomy" so that requirements are clear.
- Governance Aspects: Any legal entities with key roles (e.g. organising entity, Data Intermediary) should use the highest achievable objective in their constituent documents to retain future flexibility. Actual operations and achievement should be reviewed against these documents to make sure they are still aligned.

GGF-2.2.02/DP/F: Use Cases: What are the key Use Cases that must be enabled by the digital platform? What is the "job-to-be-done" that requires data sharing? Describe the use cases with

[58] See [33] B0.1.1 Key purpose

crisp and illustrative names: "Managing vehicle service recalls in the automotive industry" or "Calculating carbon footprint of a food product".[59]

- *Best Practice*: Define key Use Cases explicitly at the start so that requirements and objectives are clear. Use cases should have well defined actors and roles needed for success, rather than requiring broad adoption for success. E.g. Key polluters are convinced to reduce their pollution -- rather than needing to convince all polluters to reduce their pollution.
- *Governance Aspects*: Track status, success and sustainability of each use case; recognize new use cases that both re-use infrastructure capabilities as well as expanding the range of capabilities, including increasing the "objective level" (from Table 1) that the infrastructure supports.

GGF-2.2.03/DP/F: Principles, Values: Select those that will be respected in design, governance, and operation, by all members, participants.

- Governance Aspects: Should be agreed by founding members and regularly reviewed.

Consider examples of principles from other efforts:

- Staff Working Document on Common European Data Spaces[60]
  - Data Control
  - Governance
  - Respect of EU rules and values
  - Technical data infrastructure
  - Interconnection and interoperability
  - Openness
- Basic Data Infrastructure ("BDI") (Netherlands Ministry of Infrastructure and Water Management)[61]
  - Data kept at source
  - Machine-actionable systems
  - Data Sovereignty of data rights holders
  - Semantic Modelling/Metadata
  - Require Trust, but base it on a robust digital trust model
- IDSA Rulebook v2 [39]
  - Integrate existing systems
  - Integrate/use existing standards
  - Industry/domain agnostic
  - Use proven technologies
- Design Principles for Data Spaces[62]
  - Data sovereignty
  - Data level playing field

[59] See [33] B0.1.1 Use cases for data (1.1.1)

[60] European Commission. Common European Data Spaces. EC Staff Working Document, 2020. SWD(2022) 45 Final. https://digital-strategy.ec.europa.eu/en/library/staff-working-document-data-spaces

[61] BDI https://topsectorlogistiek.nl/wp-content/uploads/2022/07/20220614-BDI-Intro-FAQ-ENG.pdf

[62] Design Principles for Data Spaces https://design-principles-for-data-spaces.org/

        - Decentralised soft infrastructure
        - Public-private governance
    - EMODnet[63]
        - Collect data once and use them many times
        - Develop data standards across disciplines as well as within them
        - Process and validate data at different scales: regional, basin and pan-European
        - Build on existing efforts where data communities have already organised themselves
        - Put the user first when developing priorities and taking decisions
        - Provide statements on data ownership, accuracy, and precision
        - Sustainable funding at a European level to maximise benefit from the efforts of individual Member States
        - Free and unrestricted access to data and data products
- SITRA Rulebook for a Fair Data Economy (v2.0, Part 2) [64]:
    - Governance framework is clear, easily understood and requires minimum interpretation.
    - No pitfalls or hidden drivers/goals.
    - Transparency is expressed through documents: Content and structure based on common rulebook definition, templates and related standards. For example, adapts to regulation related to different types of data. Covers all contracts, recommendations, promises, and binding/non-binding materials like rule of conduct, including also negative use cases.
    - Governance processes address management of misuse, termination and exits (e.g., rights to data, data life cycle).
    - Control of derivative data and products is defined
    - There is a defined relationship to other rulebooks and contracts, including confidentiality agreements, as well as to existing common and domain specific laws (e.g., GDPR, IPR, health, occupational law, trade secrets, competition law, ...).
    - Contractual structure is scalable, allowing machine and distributed use, e.g., readiness to support blockchains.
    - Governance covers all participants and use cases and adheres to common laws, rules, and regulations.
    - Commitments from stakeholders and possible penalties that might be applied to stakeholders are clearly defined (e.g., Service Level Agreement; contract breach fees, trade secrets, IPR-protection, indemnification).

[63] The European Marine Observation and Data Network (EMODnet): Visions and Roles of the Gateway to Marine Data in Europe https://doi.org/10.3389/fmars.2019.00313

[64] See [33] "Clarity" section

GGF-2.2.04/DP/F: Actors Needed, Desired. Actors are specific participants that need to be involved, either for specific use cases, or overall. Are the needed and/or desired stakeholders ready for fair and trusted collaboration? Do they have limitations or minimal requirements regarding prior to joining the network? What kind of additional partners are sought for the platform? Are there other stakeholders that should be considered (e.g., officials, influencers to the data etc.)?[65].

- *Best Practice*: Many actors will be identified in Use Cases. Needed actors are usually required to perform specific roles in each use case. Desired actors represent possible participants which, if they were to take advantage of the infrastructure, their participation would signify high impact of the infrastructure.
- Links to resources needed, desired (GGF-2.2.06/DP/F): Actors that control data or services required by one or more Use Cases in turn must be identified as being needed for the infrastructure.
- *Governance Aspects*: What governance roles do needed actors need to play? Members of the organising entity and involved in the governance of that entity? Participants in the digital platform and involved in platform governance? Track and review quality of action – to measure the extent to which needed and desired actors are actively engaged.

GGF-2.2.05/DP/F: Roles Needed, Desired: What roles will the actors play?[66]

- **Best Practice**: Needed roles are defined in key use cases. Desired roles reflect possible benefits (network effects) of the infrastructure (more data providers mean more data, more data consumers mean more data available for re-use, etc.). Network effects may not be needed for the infrastructure to have impact if the use cases are successful. Roles may differ in the context of different entities/activities.
- Examples: participant, member of governance process(es), data controller, data holder, data producer, data using service, end customer, data intermediary, MyData operator, public sector actors.
- Who is in a leadership position? Are critical roles filled by actors to allow the launch of the data space?
- Note: One actor can assume multiple roles.
- **Governance Aspects**: What is the change/nomination mechanism for actors/stakeholders taking on one or more roles? Are there rules about "mutually exclusive" roles, or combinations of roles that are not allowed because of conflict of interest? Track and review quality of action – to measure the extent to which needed and desired roles are actively engaged.
- **Technical Aspects**: Several of these roles (providers and consumers) are defined in various conceptual models (Gaia-X, DSBA, MyData)

[65] References, See [33] B0.2.1 Data network stakeholders

[66] References, See [33] B0.2.2 Stakeholder roles

GGF-2.2.06/DP/F: Role definitions, expectations: Role requirements, functions, rights, responsibilities, obligations, and performance expectations, along with associated onboarding and performance monitoring processes ("rules of participation").

- **Governance Aspects**: Performance and compliance monitoring, as well as enforcement of rules of participation, must be clearly agreed and renewed by governance processes.
- **GGF-2.2.06/DP/F**: Identify classes of stakeholder and participant (as roles, possibly overlapping), with related eligibility requirements, functions, rights, responsibilities, obligations, and performance expectations in the data space. Performance monitoring processes are defined and agreed through the governance structure. Document eligibility, functions, rights and obligations, onboarding processes, performance monitoring procedures.
- **GGF-2.2.06/DP/O**: As each stakeholder and participant joins the data space, it selects one or more roles, its eligibility is determined (optionally), it agrees to related obligations and performance expectations, it is trained/oriented on how to perform its function(s), and after training, it performs the agreed functions.
- **GGF-2.2.06/DP/M-C**: Monitoring systems capture performance by each stakeholder and participant in its assigned roles; feedback is provided; role assignments (and even participation) may be modified according to agreed procedures.
- **GGF-2.2.06/DP/M-I:** All role definitions, requirements, functions, rights, obligations, and performance expectations, along with associated onboarding and performance monitoring processes, are reviewed periodically and revised as needed for the data space to meet its objectives.

As an example, the responsibilities and performance expectations for data providers include:[67]

- Making the "provided" data accessible through the platform with some level of commitment, ranging from long term to best efforts. It is the data provider's responsibility to store the data being made available unless separate repository/storage services are arranged.
- Providing the necessary "metadata" information (see, e.g., *GGF-3.1.01/DSG/F* and *GGF-3.1.06/DSG/F*) to allow the data to be used effectively, and to describe the policies that govern access and use of that data.
- Preparing the data for access and use consistent with the selected access and use policies. E.g., if data is provided without restriction, ensure the data does not include any sensitive data, including anonymization of any personal or similar data that could "poison" the data. Possibly splitting data into smaller datasets that would allow easier access to less sensitive portions of the data.
- If access and/or use of the data requires negotiation of the terms and conditions of that access, define the process for this negotiation (see *GGF-3.2.01/DSG/O+M*). Make sure proper resources are provided to support timely negotiation for access and use, regardless of whether the processes are automated or manual.
- If changes to the data or its metadata are planned, including the applicable access and use policies, as well as data structures, systems, or interfaces being used, communicating these changes in advance to varying degrees and through different mechanisms, with the most

[67] See [33] BO.4.2 Data governance; TS.3.5 Data governance solution; TS.4.2 Data location and availability.

advance notice provided to data licensees/recipients and other "concerned" participants in the data platform.

- If data will no longer be made available through the platform, what options will be available for archival access, or alternative access through other platforms?

GGF-2.2.07/DP/F: Resources Needed, Desired: Depending on the objectives of the infrastructure, various resources (data, services) may be needed/required for success, while other resources would be desirable and might improve the user experience. What data sources are required? What services are required? [68]

Examples: These examples are explored in more detail in section 3.3 since they are data space specific.

- Core services: identity management, verified credentials/presentation, attribute management, registries of participants, data sources, etc., configuration databases, service management tools,
- Reference data sets (genomes, digital cartography, vocabularies, biological taxonomies/species names, chemical databases)
- Services to help prepare data for sharing (tools that scan for copyrighted material), assist with sharing (anonymization, subsetting), creation of synthetic versions of data for safe assessment by data consumers.
- Browser services that preserve access rights (visualisation tools)
- Intermediation services, catalogues (centralised/distributed), search tools (centralised, distributed)
- Value-added services: data aggregation, fusion, enrichment, data analysis, machine learning/AI

Governance Aspects:

- Which services will be provided in the "core" (by the data intermediary, by the organising entity), which ones federated/decentralised? Are there common rules and instructions related to these services?
- The nature of the digital ecosystem must be driven by objectives and the overall business model of infrastructure -- and in turn should be supervised by governance processes.
- Data and services are provided consistent with the governance framework and specific rules and responsibilities – can this be measured and assessed to ensure compliance and performance, are periodic audits required and how will they be conducted (how often and by whom, against what standards, etc.)?

Technical Aspects: directly affects architecture, but architectural choices should be made consistent with the business model, not the other way around.

GGF-2.2.08/DP/F: Relationship between IT infrastructure (inside the Digital Ecosystem) and the Data Ecosystem. What interfaces are supported? What access policies will be needed? How can participants in the Data Ecosystem access compute and storage resources to meet their objectives? **Service "composition" and "stacking" required**. Workflows, data flows, access to and orchestration of underlying infrastructure (e.g. compute and storage)

[68] See [33] BO.2.4 Data provision; BO.5.1 Data ecosystem services; TS.4.3 Data services (technical implementation)

4.2.1.3. Generic Digital Platform: Value Creation

It is well accepted that data sharing and re-use create value as a public good. If data is "monetized" – i.e., made available for a fee or in exchange for other value, the data provider receives value for having made its data accessible and usable. However, monetization of one item of data may not be feasible: data consumers may not be willing to pay something more than the cost of preparing the data for access, advertising its availability through a catalogue and ultimately providing access through various mechanisms.

At the other extreme the altruistic benefit to a data provider for making its data freely available might not justify the effort involved. Other motivations can motivate data sharing, such as legislation (for data held by public sector bodies and generated through publicly funded research) and cultural imperatives (through expanding open science initiatives).

The digital platform must at minimum support monetization, allowing providers to set prices and providing "clearance" services that allow consumers to pay agreed fees. However, this value mechanism may not be appropriate for the use cases planned for the initiative. Ideally a data sharing platform should offer additional ways to encourage participation and engagement in the platform.

GGF-2.3.01/DP/F: Targeting, Awareness, Inclusion & Retention: How will needed and desired actors be identified, made aware of the infrastructure, and encouraged to participate? What mechanisms will be used to encourage desired participants to actively engage with the platform?

- **Design Choices**: Various economic and reputational mechanisms can be selected/defined: outreach efforts, reputational awards, branding benefits, monetary (and/or resource access) rewards.
- **Governance aspects**: Not all mechanisms will be consistent with principles/values. Governance to agree and supervise.

GGF-2.3.02/DP/F: Value sharing, exploitation, collective action, generativity: Identify mechanisms for value creation, sharing, and possibly monetization of assets/services offered through the platform. Value creation monitoring processes are defined and agreed through the governance structure. Separate mechanisms and monitoring processes may be needed for a range of value models. Document mechanisms and monitoring procedures.

How is value generated and distributed amongst the participants? What mechanisms in the platform enable value to be created through the contributions of multiple participants? Can that value then be shared with some or all of those contributing participants? Can the platform encourage collective action toward a common goal, reward the generation of new ideas/ innovation/ new ways of combining information?[69]

- How should data access or services be measured, priced and monetized?
- What kind of incentives and mechanisms are there for data sharing?

How to recognize the value of aggregated data, analysed data or the results from machine-learning models?

- Initial usage fees, subscriptions for access/use, royalties on data re-use/shared services
- Mechanisms for shared value creation and distribution of the resulting value with contributors. E.g., "kickstarter projects" for cost reductions, performance improvement, etc.
- Does the data have licensing fees or other monetary measures?
- How should data altruism organisations be handled? Should the platform sponsor such entities (e.g., a citizen science Data Altruism Organization [DAltO]?).
- Without a shared value return mechanism, volunteers may be discouraged from bringing their best ideas if there is insufficient return to them, and/or too much free ridership.

If created value is embodied in new data or a new service, is it possible to mark these resources as "club goods", and give contributors privileged access to such resources?[70]

Governance aspects:

- Track concerns with "tragedy of the commons", "free riders", etc.
- Supervise return of shared value to contributors, provide appeals mechanisms in case of disputes.
- Supervise access limits on club goods
- What are the safeguards and monitoring mechanisms for value?

GGF-2.3.02/DP/O: As assets/services are offered through the platform, value-creation events are recorded and tracked.

GGF-2.3.02/DP/M-C: Value creation is monitored, reported and evaluated; feedback is provided as appropriate; value-creation mechanisms and related monitoring processes may be modified as needed for the data space to meet its objectives.

[69] See [33] B0.3.2 Data value

[70] The potential value of club goods in connection with environmentally-related data is discussed in Fritzenkotter, et al. "Club goods" are defined at https://en.wikipedia.org/wiki/Club_good.

GGF-2.3.02/DP/M-I: Governance processes around value creation and sharing, as well as the mechanisms included are assessed to find opportunities for improvement and innovation.
GGF-2.3.03/DP/F: User Experience: Ensure each user's experience with the infrastructure helps achieve the objectives of the platform, contributes to the value that can be generated, and does not detract from either. The user experience should be consciously designed and assessed regularly for effectiveness. • For example, data.europa.eu assessed its user experience and made 10 recommendations for improvement[71]. All recommendations impacted the design of the portal, but most also required the collection of additional information on each item of data (e.g., ethical, statistical) that translated into new metadata requirements.

4.2.1.4. Generic Digital Platform: Technical Architecture and Control

As a "digital infrastructure" designed to enable "data transactions" between "ecosystem parties", it is clear that decisions must be made about the design and implementation of the data platform. Many of the governance requirements identified throughout this document will determine specific functionality and features that must be supported. However, these technical decisions should follow agreement among stakeholders about the "shape" of the technical solution, which in turn should align with the Landscape Design, Mission, Actors/Resources/Roles and Business Model decisions addressed earlier. Note that the scope of technical architecture includes software, software development, technical standards, as well as a range of "boundary resources" that range from standard documentation and training to more elaborate mechanisms for outreach to and support of technical personnel (working "inside" the data platform as well as with and for various participants such as data providers and consumers).

Control refers to the ongoing governance of the technical architecture, including technical planning, roadmap maintenance, change management and technical innovation more generally. Technical change should be purposeful – consistent with the objectives of the data space and its stakeholders – and support achievement of progressively higher objectives (see Table 1).

[71] European Commission. "Principles and recommendations to make data.europa.eu data more reusable". April 2022. https://data.europa.eu/sites/default/files/report/D3-4-1-1-Strategy-Mapping-Report-v3-0.pdf

GGF-2.4.01/DP/F: Architecture, Control: Comprises the modular architecture within the boundaries of the infrastructure, the definition of its internal interfaces and the compatibility to relevant external systems.

Technology aspects[72]:

- What are the design principles, focus areas and design philosophy for the platform's common technology solution?
- What existing data sharing, infrastructure and other reference solutions are used as the basis for the common solution? (IDS Connectors, Eclipse Data Connector, other interfaces? Catalogue standards (DCAT, geoDCAT). Planning for new capabilities such as SIMPL)
- What are the key functional and non-functional requirements, available standards and reference implementations, interfaces and APIs, common roadmap?
- What kinds of interfaces does the solution provide?
- What interface descriptions are needed? How are they defined?
- How mature are those interfaces or are changes expected?

Governance aspects:

- Design choice: Creating a "core/exchange" architecture that encourages creation and availability of value-added services, that encourages creation of new high value data.
- How will technical decisions be managed? How will these decisions be aligned with the objectives of the platform and relationships with the broader landscape?
- What are the key decisions related to overall architecture and technology choices (e.g., cloud solution, vendor independence)?
- How will the evolution of the interfaces be managed, e.g., in regard to backward compatibility? What is the plan and/or roadmap for their evolution?

GGF-2.4.02/DP/F: Openness, Modularity, Intellectual Property Policy: Openness refers to "the easing of restrictions on the use, development and commercialization of a technology", contrasting with closed, proprietary approaches. Practically this is implemented through the "Intellectual Property Policy" of the platform, identifying the various points in platform operation where intellectual property (IP) may be created, and identifying the corresponding intellectual property rights (IPR) held by related roles.

- Develop and approve an Intellectual Property Policy, identifying situations where IP might be created, the roles that might be involved in its creation, and the IP rights that may be available to involved participants based on their role(s). The IP Policy will also define the process for declaring existing IP that might be made available to the platform, for reporting newly created IP and the participants involved in its creation.

GGF-2.4.02/DP/O: Manage the IP declaration and reporting process, the identification of new IP and corresponding IP rights that have been agreed.
GGF-2.4.02/DP/M-C: Monitor and report the overall flow of IP in the platform, including the effectiveness of the process for reporting new IP and defining new IP rights.
GGF-2.4.02/DP/M-I: Monitor the operation of the Intellectual Property Policy overall. Revise through the associated governance process, as needed to better meet the objectives of the platform.

[72] See [33] TS.2.1 System design principles; TS.3.1 Technical interfaces

For example, analysis capabilities might be offered using free open-source software, with open-source licences from the original developers. Data space participants could "adapt" the algorithms for new hardware or software environments or with additional language support, but the original developer must be given the rights to this adaptation. Conversely, participants can "extend" the capabilities and functions of the software, in which case the extending developer and the original developer will share the IPR in the extended version of the software.

GGF-2.4.03/DP/F: Centralised or Distributed: Within the data sharing infrastructure, alternative architectural approaches can be used to deliver the services and functionality required. The selected architectural approach must be defined at the formation stage. Every required service and/or function must be provided by one or more service providers, raising governance questions about how those providers are selected and paid, how the quality of their services is supervised, and how the interactions between them are coordinated.

- A "centralised" service is one that is not distributed among multiple providers, such as management of a registry, e.g., of participants, resources, and trusted identity providers.[73]Does the organising entity intend to operate these services, either initially ("bootstrapping") or on a continuous basis? Is the organising entity capable of this activity (e.g., if the organising entity is NOT a legal entity, who will take legal responsibility for the services and any liabilities associated with their provision?). Can the organising entity arrange to outsource this service to another organisation? Does the business model (see **GGF- 2.4.10/DP/F** below) provide revenues to cover the cost of providing these central services?
- Distributed services can be performed by multiple providers in both a collaborative or competitive way. Federated identity providers operate in this way, as do distributed computational services (cloud, grid computing). Governance processes are required to coordinate these providers, ensure both technical and operational compliance with agreed standards, and manage the onboarding and possible removal of providers according to agreed rules and procedures. Such services can either be paid for as "core" services of the platform, covered by the platform's business model, or on a pay as you go basis as resources are consumed.
- For both centralised and distributed services, the organising entity can delegate required services to one or more providers, but the organising entity would need to define the scope of each service, select one or more competent providers for each service, and manage the services provided through a service management system (see **GGF-2.4.09/DP/F**).

GGF-2.4.04/DP/F: Participant/User Identification, Registration, Trust Framework: A trust framework enables relationships between and among the organising entity and participants, such as service and software providers, as well as for the relationships among and between

[73] Such registry services can also be provided in a fully decentralised way using blockchain-based architectures. However the interaction between architecture and governance in such systems is complex and beyond the scope of this analysis, and as noted in section 1.4 above, Chapter 4, we exclude fully decentralised infrastructures from our discussion.

platform participants more generally. The infrastructure's trust framework will specify how participants are identified digitally, including whether and to what extent anonymous users might be able to use any of the services available through the platform. This may include registration requirements.

Technical aspects[74]:

- What is the solution for identity, roles and access control?
- Trusted identification of platform participants? How are identities created and governed?
- Are there additional requirements for the identity and access management not readily solved by the selected solution, such as data stream identities or need to merge or split identities?

Governance aspects:

- Should the scope of the trust framework be established across multiple digital ecosystems, across multiple data space initiatives within a data ecosystem, for each individual data space?
- If data providers will not accept technical controls on data use (such as secure multi-party computation or other privacy preserving techniques), can the trust framework be designed to encourage providers to make sensitive or valuable data available?
- Can provisions be added to various legal agreements (e.g., in the Constitutive Agreement) that would support and strengthen the trust framework?
- Should digital identities be required to match, or "bind", with real legal or natural persons, to link legal responsibilities with responsibilities assigned and assumed in the data ecosystem? Does this consideration drive the scope of the trust framework?

[74] See [33] TS.3.2 Access control and Identities.

GF 2.4.05/DP/F: Cybersecurity: If cybersecurity is required for the data sharing infrastructure, define the structures needed to fulfil this requirement:

Security risk and threat assessment[75]

- How are the risks and threats identified and assessed?
- Security risk assessments need to consider not only physical security and individual organisational issues, but also the risks associated with networks and network interoperability.
- How are the risks at the data platform level collectively identified?

Data and data network related threats[76]

- What threats are related to data and the operation of the data platform?
- What general data security threats should be addressed in the governance framework? How to manage and prevent potential challenges in the data platform and services related to it? These threats include unintentional or intentional disclosure of data, user-based threats (phishing, social manipulation, access control), data hijacking (man-in-the-middle), insider threats, and technical threats such as data loss, ransomware, and cloud challenges.
- Which data-related security threats should be addressed in the governance framework? These threats include misuse of data, data leaks, inaccurate or poor quality of data, and data-related liability issues.
- What threats are related to data and the operation of the data platform?
- What general data security threats should be addressed in the governance framework? How to manage and prevent potential challenges in the data platform and services related to it? These threats include unintentional or intentional disclosure of data, user-based threats (phishing, social manipulation, access control), data hijacking (man-in-the-middle), insider threats, and technical threats such as data loss, ransomware, and cloud challenges.
- Which data-related security threats should be addressed in the governance framework? These threats include misuse of data, data leaks, inaccurate or poor quality of data, and data-related liability issues.

Security objectives and regulation[77]

- What are the security objectives of each participant and the data network as a whole?
- Do specific regulations address the data security of the planned data platform? Security objectives should be defined from the perspective of both the individual participants and the platform as a whole.
- How has security been addressed in existing data sharing solutions? What is the existing legislation in this area?

Risk and security management process and tools[78]

[75] See [33] TS.5.1 Security risk and threat assessment.
[76] See [33] TS.5.2 Data and data network related threats.
[77] See [33] TS.5.3 Security objectives and regulation.
[78] See [33] TS.5.4 Risk and security management process and tools.

- Identify the risk and security management process and tools that are required for the data platform. Include them in the Service Management System (see **GGF-2.4.09/DP/F**).
- Once threats and vulnerabilities have been identified, the severity of the threats to the data platform can be assessed, for example by determining the probability of each risk and the magnitude of the damage if the risk materialises. This will help identify the risks that are most critical to address in the design of the data platform.
- Include expected exception management and damage control requirements in the Service Management System
- What combination of management tools will achieve the required level of security and transparency?

Confidentiality of data[79]

- How is confidentiality of data defined and managed in the data platform? What is the value of information to the various parties involved?
- What is the damage if information is intentionally or unintentionally disclosed to 3rd parties without the consent of the provider, or if it is used in breach of contract?

[79] See [33] TS.5.5 Confidentiality of data

GGF-2.4.07/DP/F: Boundary Resources: Identify and create required boundary resources, as well as related performance and satisfaction measures. Boundary resources are tools, regulations or other resources that govern co-creation of value in platform ecosystems. Bianco et al. (2014)[80] further differentiate boundary resources into:

- *Application Boundary Resources*: technical resources that enable services or components to interact (operationally or at "run time") with the other services or components in a digital platform. Application boundary resources include APIs, technical and data standards.
- *Development Boundary Resources*: technical resources or tools that enable services or components to be developed, adopted, and maintained within the digital platform. Development boundary resources could be extensive, focussing on creators of new components, including software development kits (SDKs), development environments (IDEs), debuggers, test suites and certification processes, or more straightforward, focusing on enabling adoption of the platform by new participants, with repositories of approved components that can be tested in a sandbox environment and then deployed in production after acceptance by the participant.
- *Social Boundary Resources:* Understanding and working with both types of boundary resources requires specific knowledge transfer about how these resources work,typically in the form of training material, promotion and training events, and online community forums. Additional resources could include incentives for adoption, guidance on copyright and intellectual property policies, as well as operational guidelines and documentation for the digital platform. Accessing Social Boundary Resources is as important to platform participation as the availability of Application and Development Boundary Resources.

GGF-2.4.07/DP/O: Deliver and maintain required boundary resources.

GGF-2.4.07/DP/M-I: Monitor the effectiveness of boundary resources in helping stakeholders and participants perform required functions. Identify needs for new or improved boundary resources

GGF-2.4.08/DP/F: Development Plan: Plans for the ongoing development, growth of, and innovation within, the contemplated data sharing infrastructure should be created during formation, to ensure a holistic design and comprehensive governance.

GGF-2.4.09/DP/F: Service Management System: The data platform provides services to its participants, and these services must be managed to ensure the platform operates effectively, meeting or exceeding the expectations of "customers" and meeting overall objectives.[81] A service management system (such as FitSM[82]) should be selected at the digital platform formation stage, including a range of processes and capabilities to ensure effective service delivery, such as:

[80] V. D. Bianco, V. Myllärniemi, M. Komssi and M. Raatikainen, "The Role of Platform Boundary Resources in Software Ecosystems: A Case Study," *2014 IEEE/IFIP Conference on Software Architecture*, Sydney, NSW, 2014, pp. 11-20, https://doi.org/10.1109/WICSA.2014.41.

[81] See [33] Bo.4.3 Risk management

[82] www.fitsm.eu

- Incident and problem management. How are incidents or disputes (both service and data-related) managed?
- Risk identification, management and mitigation processes.
- Technical change management processes[83]

GGF- 2.4.10/DP/F: Business Case and Model: What is the business case for the platform? What is the business model? [84]

- Costs: Include costs for development and operation of the platform: technical implementation, creation of boundary resources, maintenance of assets, service management, communications, outreach, training, governance costs, management costs, monitoring of performance and compliance, support for participants in using and exploiting the platform. In designing the business model, consider "make", "buy" and "outsource/ contract" development and operation options.
- Revenues: How will these costs be recovered – member contributions, participation fees, transaction fees, advertising, etc.? Will cost recovery be different for different participants? Will there be volume discounts?
- Performance and Sustainability: How to ensure continuity of joint operations? For example, how do we ensure both fair use and fair supply of data in the network? What kind of strategic bi-directional dependencies exist between the partners of the data network?[85]

4.2.1.5. Generic Digital Platform: Governance Architecture

In parallel with the design of the platform's technical architecture, specific governance mechanisms must be designed to ensure that the platform supports the needs and objectives of its stakeholders, fulfilling its agreed mission and vision, and remains relevant, effective and sustainable.

The process of identifying the actors and resources required and desired for the digital ecosystem and its digital platform is the first step in defining the governance architecture. Required actors, and parties responsible for required resources, including financial resources, define the primary list of stakeholders that may need to be included directly or indirectly in governance. Other stakeholders include those accountable for the results that might be enabled by the ecosystem. In the case of the European Green Deal, this might include different levels of government, civil society, industry or other stakeholders that might benefit from, or might be negatively affected, by the actionable insights, target setting, or other outcomes enabled by the Green Deal Data Space.

GGF-2.5.01/DP/F: Identify Possible Entities that Must be Governed: These could include the Digital Ecosystem as a Community of Practice, the Digital Platform as a technical infrastructure, the Organising Entity of the Digital Ecosystem (a legal entity), Data Intermediaries (legal entities),

[83] See [33] TS.4.1 Change control.

[84] See [33] B0.3.1 Business case; B0.3.3 Data network solution fundamentals.

[85] See [33] B0.3.4 Level of commitment.

Collective Data Providers (such as a Data Altruism Organisation for Citizen Generated Data), etc.

GGF-2.5.02/DP/F: Map Stakeholders to Governance Roles: Roles can include formal "membership" as a voting member of one or more legal entities, informal "voting rights" in the governance of communities of practice and technical platforms, and advisory roles at several levels.

- Although specific actors may be solicited to take on governance roles, rules for participation in governance ("membership rules") should be defined more generically so that they can be used to manage membership on an ongoing basis.

GGF-2.5.03/DP/F: Define Governance Bodies for Each Entity, related Authority Levels, Decision-Making Mechanisms: Traditional governance bodies fall into three categories:

- **General Assembly**: All Members of an Entity. The Authority of the General Assembly may be limited to electing the Governing Board, approving annual financial reports, and approving changes to the formative agreements of the entity (such as the Constitutive Agreement and/or By-Laws of a legal entity). Some General Assemblies have broader powers, including approving strategy for the organisation, annual budgets, expenditures over a certain amount, etc.
    - **Decision-making mechanisms**: Depending on the decision to be made, different decision-making processes may be required, such as majority or supermajority voting, advance notice of matters coming up for a vote, quorum requirements.
- **Governing Board**: a smaller group entrusted with key decision-making powers for the entity. The governing board typically has authority to review and approve all aspects of the entity's operation, except those powers held by the General Assembly. The governing board has the responsibility to make operational arrangements for the entity under its supervision, but typically delegates many of those responsibilities and authorities to an executive manager, whose plans and performance the board then supervises. All of the topics itemised in this deliverable require decision-making by the governance body of the relevant entity.
    - Governing Boards can be designed to include representatives of key stakeholders, individuals experienced in governance of similar activities, as well as experts in topics of strategic importance to the entity.
    - Larger Governing Boards may delegate some of their authority to executive boards who can act on critical issues in a timelier fashion than a larger board.
    - Decision-making mechanisms need to be defined (see list above for possible mechanisms).
    - Governing Boards may have advisory boards or committees to guide their decision-making. Advisory boards can provide representation from broader groups of stakeholders. Committees can be set up to address specific administrative matters.
- **Executives and management**: While executives, and management more broadly, are not involved in governance, their jobs begin where governance ends, so interactions between management and governance processes should be tracked transparently. Governance actions assigned to management are tracked and reported, along with results. Management will report to governance bodies about performance of the entity and is held accountable for those results.

GGF-2.5.04/DP/F Define Member Management Processes: Define Membership rules of participation, onboarding process, management of member rolls and audits of compliance with rules of participation.

GGF-2.5.05/DP/F: Define Governance Tracking Processes: This includes: ● Notices of meetings, advanced agendas, minute-taking in meetings, open and *in camera* deliberations. ● Formal consultation processes, properly identifying stakeholders, issues being consulted, consultation time periods, alternative modes of consultation (town halls, surveys, etc.) Follow-ups to decisions: status of implementation, measurement of results, etc.
GGF-2.5.06/DP/F: Define Legal Structures: Set up structures for governance: ● For each legal governance entity, define its legal form and jurisdiction. ● For each informal governance entity, adopt a similar structure to ensure consistent and well-defined procedures. ● Engage competent advisors to create appropriate legal structures. ● Structures should reflect decisions above about what entities are needed, who should be involved in governance, the mechanisms for this involvement, membership rules and process, how governance decisions are tracked and communicated.
GGF-2.5.06/DP/GUIDANCE: Develop best practices and templates for legal agreements forming a data sharing infrastructure. Consider any provisions of the Data Act that may make such "standard" agreements unenforceable with SMEs, since normally they would not be open to negotiation.
GGF-2.5.07/DP/F: Pre-launch consensus building: Working Group of key actors to develop mission, scope, governance (membership and decision-making, governance rules, risk management), platform design, key actors, business model, funding, operating plan (communications, onboarding, security, innovation, training), roadmap

4.2.2. Generic Digital Platform: Formation – Launch Milestone

For a new data space, the transition from the Formation stage to the Operation and Monitoring stage represents a key milestone – the "launch", involving several specific activities that are critical to the ongoing success of the data space. These activities touch most aspects of governance, so they are presented here.

GGF-2.6.01/DP/F-L: Create Final Founders' agreements, Constitutive Agreement: Documentation package, reflecting agreed consensus.
GGF-2.6.02/DP/F-L: Incorporate: Multiple levels and entities may be appropriate (e.g., EDIC, separate Data Intermediary entity, core service providers, citizen science, Data Altruism Organisations)
GGF-2.6.03/DP/F-L: Initial Member Onboarding: Founding members are replaced by a broader group of members onboarded through an agreed process.
GGF-2.6.04/DP/F-L: Formation of Legal Governance Bodies and Processes: Boards and advisory boards/committees are formed following agreed structures.
GGF-2.6.05/DP/F-L: Registration with competent authority: If a legal entity has been created, among other things, to act as a Data Intermediary (under the Data Governance Act), the entity must register with the appropriate competent authority as specified in the DGA.
GGF-2.6.06/DP/F-L: Formation of Operational Governance Bodies and Processes: For the operation of one or more data sharing initiatives within the scope of corresponding legal organising entities, operational boards and advisory boards/committees are formed following agreed structures.
GGF- 2.6.07/DP/F-L: Pre-and-post launch outreach/communications to stakeholders: Ensure stakeholders are kept informed and engaged in the efforts of the Platform.

GGF- 2.6.08/DP/F-L: Co-design technical development process: Recognizing that "off the shelf" data space technologies are not available, start the DevOps (development plus operations) activity needed to identify and assemble an interoperable suite of tools that can be reliably deployed to support multiple data spaces/data space initiatives.

4.2.3. Generic Digital Platform: Operations and Monitoring

As noted above, most initial design decisions are complemented by the need for operational processes to implement those decisions. In parallel with actual operations, those operations should be monitored to ensure that the data sharing initiative complies with external requirements as well as the intent of the initiative's own governance processes. In addition to compliance management, activities should include regular risk management, as well as reflection to identify any opportunities for performance improvement.

Here we complement any operations and monitoring requirements identified above with notable operational requirements that must be supported.

GGF-2.7.01/DP/O&M: Management of Members (participants in governance processes): Membership rules of participation, onboarding, management of member rolls and audits of compliance with rules of participation.
GGF-2.7.02/DP/O&M: Decision-making: Decision-tracking, consultation, governance body interactions. This includes: ● Notices of meetings, advanced agendas, minute-taking in meetings, open and *in camera* deliberations. ● Formal consultation processes, properly identifying stakeholders, issues being consulted, consultation time periods, alternative modes of consultation (town halls, surveys, etc.) ● Follow-ups to decisions: status of implementation, measurement of results, etc.
GGF-2.7.03/DP/O&M: Communications: Communication of decisions to stakeholders, measurement of effectiveness of the communication strategy/plan, and channels. Including: ● Analyse and gather feedback for improvement. ● Tailor communication to target users. ● Ensure user consent and preferences are respected for the different communications channels.
GGF-2.7.04/DP/O&M: Onboarding Participants: Application of agreed inclusion criteria are likely to specify alignment of the participant with the principals and objectives of the data space initiative, commitment to quality and sustainability, other assessments. ● Technical aspects: Ideally onboarding would be automated, but many dimensions of the onboarding of participants may be subjective. Automated onboarding may be possible, e.g. using verified credentials that would confirm facts about a participant such as legal status and jurisdictions of establishment, or successful certifications. Gaia-X and the DSBA present structures to support such approaches, but not every inclusion criterion can be automated.
GGF-2.7.05/DP/O&M: Onboarding Services & Data: Application of agreed Inclusion criteria, such as fitness for purpose as measured by quality standards, completeness of metadata annotations, declared levels of sustainability (i.e., for how long in the future will the data provider commit to updating the data), etc. ● Technical aspects: Automated onboarding could be important for services and data since their volume will be greater than for data providers *per se*.

GGF-2.7.06/DP/O&M: Development Activities: Technical, as well as documentation, training materials, support services, etc.
GGF-2.7.07/DP/O&M: Cybersecurity Activities: Involving People, processes, technology as well as physical security.
GGF-2.7.08/DP/O&M: Innovation & Growth: Build processes that advance the maturity and capabilities of the people and systems involved in the data spaces/data space initiatives. Assess current skills sets, identify gaps and opportunities, act to fill those gaps and exploit those opportunities.
GGF-2.7.09/DP/O&M: Training: e.g., improving data awareness and data literacy among stakeholders, increasing skills of stakeholders to engage with one or more data spaces/data space initiatives, working with boundary resources and actively participating in data transactions.

As noted above, monitoring activities cover several specific operational processes, and in general monitoring should support the governance process with a comprehensive view of the operation of the initiative. These comprehensive views are provided by the following processes:

GGF- 2.7.10/DP/O&M: Service Management: Employ a structured Service Management System to measure and manage delivery of services reliability and efficiently, achieving high satisfaction among stakeholders.
GGF-2.7.11/DP/O&M: Compliance Monitoring and Management: Operate compliance monitoring systems that measure compliance with agreed policies, rules and regulations established for the data space, as well as with relevant legislation (EU and member state, horizontal and sector-specific).
GGF-2.7.12/DP/O&M: Performance Monitoring and Management: Measure performance against agreed indicators (KPIs), monitor trends, identify opportunities for improvement or emerging risks.
GGF-2.7.13/DP/O&M: Risk Monitoring and Management: Create and maintain a register of risks related to legal compliance, stakeholder satisfaction, achievement of overall objectives. Evaluate risks to identify those likely to create significant impact on individual entities and on the data sharing enterprise more generally.

4.2.4. Generic Digital Platform Sustainability

Regardless of the agreed business model for this data space (see *GGF-2.4.10/DP/F: Business Case and Model*), sustainability cannot be taken for granted.

- If the data space's business model is expected to achieve sustainability through growth of one or more metrics (e.g., number of data transactions, participation in shared value created through use cases, number of participants through participant or onboarding fees), these metrics need to be measured and progress toward sustainability tracked.
- If the business model requires external investment from partners and/or funding bodies, the enterprise must align itself with the objectives of those partners and funding bodies,

gather data to show how their objectives are being achieved (e.g., through measured impacts, success stories, satisfaction surveys, etc.).

GGF-2.8.01/DP/S: Financial Sustainability of the Digital Platform: Based on the agreed Business Model, establish ongoing processes for monitoring the performance of the data space enterprise in terms of measures of its sustainability (e.g., financial reports, impact reports, etc.). Resource initiatives to improve financial performance, capture/improve impacts, etc. as needed to achieve the sustainability objectives, including updating business plans and reporting, preparing funding proposals, etc.

In the specific case of "data intermediaries" as defined by the Data Governance Act, these entities are required by the DGA to arrange for the continuity of their provided "data intermediation services".

GGF-2.8.02/DP/GUIDANCE: Interpretation of the DGA's Continuity Requirement for Data Intermediation Services: The scope of this requirement is unclear. Is continuity required under all circumstances or only where the Data Intermediary is storing data on behalf of Data Providers. For how long is continuity required, or is it acceptable to arrange for another entity to take over either the data intermediation service or just the data storage service?
GGF-2.8.02/DP/S: Continuity of Data Intermediation Services: Negotiate and maintain required continuity for any Data Intermediation Services to comply with DGA requirements

4.3.Generic Digital Space Governance

As the concept of data spaces has gained acceptance and moved from idea to implementation, several analyses have considered how such data spaces should be governed and managed to ensure their objectives are met.

Many of the identified issues have been addressed in the preceding section (4.2) in the context of digital platform governance. This section addresses the governance framework specifically required to encourage and enable "data transactions", while the next section (4.4) explores the governance framework needed to respect requirements set by data holders on the acceptable visibility, access and use of the data they hold.

As in the section above, the sections below follow the design/governance lifecycle to present the key dimensions of data space governance that must be considered in the creation of a new data space.

4.3.1. Generic Data Space Governance: Formation

At the formation stage for a data space, it is important to establish a clear structure, or taxonomy, to model how services are offered to users, how they might be combined, and, which ones are wholly or partially addressed by the Data Governance Act (DGA) and therefore need to be managed and governed in a specific way. (See Section 4.1, Governance Requirement *GGF-1.03/DSG/F*).

GGF-3.1.01/DS/F: Domain Data Models, Interoperability Standards: Define/select the data models and interoperability standards required by the domain to ensure semantic interoperability. The Data Act[86] requires that data providers must describe the "data structures, formats, vocabularies, classification schemes, taxonomies and code lists … in a publicly available and consistent manner."

Technical aspects[87]:

- Identify models and model transformations that need to be supported for effective use of the data.
- What is the format and structure of data and associated metadata? Is this structure described and shared?
- What data standards are used?
- Are data models semantically compatible? Are differences significant? How are the incompatibilities resolved?

Governance aspects: Define a governance process for consideration of, and consultation about, new or changed models, decisions about their adoption, and processes for implementation of changes by the data space and by the involved providers.

GGF-3.1.02/DS/F: Intermediation/ Marketplace/ Catalogues: What services are needed to enable data transactions?

Governance Aspects*:* Supervising categorization of data intermediation services, monitoring performance and compliance.

Technical Aspects*:*

- Possible functions for catalogues are outlined in "Data Catalogues" — Implementing Capabilities for Data Curation, Data Enablement and Regulatory Compliance - 2022 Edition"[88]
- Various technical services are described as Technical Building Blocks, as well as technical convergence work -- which ones are Data Intermediation services (DIS)? Those that are DIS need to be "instrumented" to collect data required for compliance. Can DIS be delivered on a decentralised basis?
- If so: Can Data Intermediary treatment be avoided with distributed/decentralised architecture?

GGF-3.1.03/DS/F: Ancillary Services: Data Preparation, Encryption, Anonymization, Transformation. What services make intermediation easier/more likely?

- Examples: data preparation, repository, format conversion, compliance assessment
- *Technical aspects*: Various technical services are described as BBs, as well as technical convergence work -- some fall into this category of ancillary data services.

[86] Data Act Article 28

[87] See [33] TS.2.2 Metadata and data formats (4.2.1, 4.2.2

[88] https://www.isst.fraunhofer.de/content/dam/isst-neu/documents/Publikationen/Datenwirtschaft/Fraunhofer-ISST_DataCatalogs_Report-kl.pdf

- Anonymization Services should only be offered if they can be relied upon to convert "personal" data to non-personal data that is not subject to the GDPR. Otherwise, data providers should take responsibility for the provision of anonymized data that is not covered by GDPR.

GGF-3.1.04/DS/F: Enrichment, Aggregation, Fusion, Analysis, AI/ML. What services are needed or valuable AFTER data is shared?

Governance aspects: Supervising onboarding of value-added services, monitoring usage, performance (value created), principles (value sharing), compliance. Rules for access to data results by contributors? Rules against bundling of "interesting" private sector data with costly data fusion services? Should governance have the "ability" to review services with respect to principles/values?

Technical aspects: In general, how should value added services be "instrumented" to enable compliance with the AI Act?

GGF-3.1.05/DS/F: Forecasting, Monitoring, Trend evaluation, Target setting and tracking, alerting, dashboarding. Services needed to enable higher levels of "Purpose".

GGF- 3.1.06/DS/F: Creating a Framework for Declaring Access, Usage Policies. What are the permissions and restrictions on data use?[89]

- The Data Act[90] requires that data providers must specify the "use restrictions [and] licences […] sufficiently […] in a machine-readable format" to allow a prospective data consumer to find, access and use the data. The Data Act also requires that data providers must provide "the means to enable interoperability of contracts for data sharing."
- Access and use policies should be consistent regardless of the technical access mechanisms that may be required – e.g., direct data download from data providers, indirect linking to a repository employed by the data provider, or "data as a service" mechanisms for accessing large or complex data sets (e.g., Copernicus earth observation data). This consistent approach should also apply to situations requiring "data visiting", "compute to data" or secure processing environments.
    - The Data Act91 requires that data providers must specify "the technical means to access the data, such as APIs and their terms of use and quality of service shall be sufficiently described to enable automatic access and transmission of data between parties, including continuously or in real-time in a machine-readable format where that is technically feasible and does not hamper the good functioning of the product."

[89] See [33] B0.5.2 Data usage control

[90] Data Act Article 28. The specifics of this Article may be supplemented and further defined by delegated acts developed in consultation with the EDIB.

[91] Data Act Article 28. The specifics of this Article may be supplemented and further defined by delegated acts developed in consultation with the EDIB.

- **Standard contractual terms.** A data space may establish standard terms and conditions for data transactions as part of its governance framework. Some terms and conditions may allow for variation according to the needs of data providers and consumers, while other terms and conditions are expected to be standard for all data transactions enabled by the data space. Standard terms might include:
    - Participant rights and responsibilities
    - Audit rights
    - Applicable law and dispute resolution
- Forms of licences and terms of licences
- Human-readable and machine-readable

Technical Aspects: A number of "rights expression languages" (RELs) have been proposed, including Extensible Access Control Markup Language (XACML)[92] and the Open Digital Rights Language (ODRL)[93]. These RELs can be translated to other REL formats[94], however the semantics of specific terms in each REL need to be aligned and mapped to real-world actions and activities involving the specified assets.

[92] eXtensible Access Control Markup Language (XACML) V2.0, http://www.oasis-open.org/specs/index.php#xacmlv2.0
[93] http://odrl.net
[94] Maroñas, Xavier & Rodriguez, Eva & Delgado, Jaime. (2023). An architecture for the interoperability between Rights Expression Languages based on XACML.

- Specifying disclaimers or limitations of liability for potential data accessors/users. Are these defined in a machine-readable way?
- Is exclusive access and/or use of data possible? Can the possible uses of the data be limited geographically or by economic sector? E.g., A data consumer might seek to exclusively access and use data to create value-added data to be provided to the consumer's own customers in specified countries and/or specific economic sectors (e.g., regulated banks in the EU). Such exclusivity might be agreed by the data provider for a defined period (e.g., X years), after which access to the data might be renegotiated or potentially offered to other consumers.
- Confidentiality and/or use of the data limited to internal use by the consumer.
- Rights of, or limitations on, users creating new Intellectual Property through the use or manipulation of the data. Licences might require the data consumer to remunerate the data provider through a form of royalty on any new intellectual property derived from the data.
- Provisions for management of conflicts between different users of the same data, e.g., if exclusive rights are given to two different users, and the rights are found to overlap (e.g. "banks" vs. "financial institutions").
- Defining other liabilities related to real-world actions and activities involving the data.
- Defining role of possible 3rd parties. Do usages policies allow transfer or distribution of data to 3rd parties? If so, how can 3rd parties use the data? What are they allowed to do, what permissions, prohibitions and obligations apply to each 3rd party? Which party is responsible for 3rd party infringements of agreed policies?
- Penalties and remedies for violations of specified usage policies.
- Mechanisms for negotiating access and usage policies. (See Contract Negotiation *GGF-3.2.01/DS/O+M*.)
- Mechanisms for ensuring protected classes of data consumers (e.g. small and micro enterprises) can avoid "unfair" terms of access and use.

- Usage policies may include the 15 patterns defined by IDSA[95], which fall into the following categories:
  - Generically allow/prohibit a category of usage (referring to actions such as "use", "read", "distribute", "print", etc., although these actions are not further defined) (#1)
  - Limitations on how data may be accessed, such as:
    - specific IDS Connectors (#2),
    - specific IT systems or applications (#3),
    - by specified users or groups of users (#4),
    - at specific locations (#5),
    - for specific purposes (#6),
    - after occurrence of a specific event (#7),
    - IDS Connectors with a specified security level (#8).
  - Limitations on time of access (time interval (#9), time duration (#10), number of "uses" (#11))
  - Use and delete (#12)
  - Modify data in transit (#13) or at rest (#14)
  - Log data usage (#15), notifications when data is used (#16), attach policy(ies) when distributing to 3rd parties (#17), distribute only if encrypted (#18)
  - Perpetual licence (#19)
  - Licence subject to subscription payment (#20)
  - Use subject to external conditions (e.g. status of data exchange technology, contract terms, etc.) (#21)
- "Type" of data: Does the data contain photos, audio – or video content, computer programs, etc. that have special legal requirements? [96]
- *Sui generis* databases: Are database rights applicable to data (i.e., data has been collected and organised into a distinct entity (the database) entailing substantial effort on the part of the creator)? Are single records from such databases offered as data – is the legal prohibition on "extraction" being waived? [97]
- Limitations on proposed purpose of use, e.g.:
  - Research
  - Response or preparation for emergency by public sector body
  - Training of AI models
- Limitations on categories of actors which might be given access and the ability to use, e.g.:
  - any Participant in this Data Space,
  - only Participants which the Data Provider has qualified as not being a competitor or the Provider,
  - only Participants with certain certifications (e.g. ISO 27001, ISO 9001) or other documented qualifications, etc.
- Limitations on where data may be accessed:
  - Geographic limitations (data can only be transferred to certain jurisdictions (whitelist), or certain jurisdictions are unacceptable as destinations (blacklist))
  - Infrastructure requirements and limitations (destination infrastructures and/or network infrastructure with certain documented qualifications or certifications, e.g. operated by ISO 27001 certified entities, Trusted Execution Environments, etc.),

GGF-3.1.07/DS/F: Define the Process for Ordering Data or Requesting Data from "Data as a Service" Services.
GGF-3.1.08/DS/F: Identify supported mechanisms for Data Transfer: Does this occur inside or outside the data space? Through what mechanisms and protocols? Are secure networking facilities required?
GGF-3.1.09/DS/F: Transaction Logging and Usage Accounting. How will data transactions be recorded by the data space, and how can these logs be reviewed and analysed to ensure compliance and measure performance?[98] **Technical aspects:** ● How will data transactions be monitored? ● What capabilities are required? ● Agreeing and confirming transactions, e.g., digital signatures, access keys and identities? ● Monitoring and reporting of system and data use (e.g., monitoring APIs)? ● What additional information needs to be collected to ensure compliance when different data types are exchanged e.g. personal data?[99] **Governance aspects:** ● How will the performance and compliance of data transactions be monitored and governed? ● Agreeing on the scope of monitoring and transaction logging, and the ability to track user actions for accounting and compliance. ● How are permissions for personal data processing technically managed, logged, monitored and reported?

4.3.2. Generic Data Space Governance: Operation and Monitoring

Most operations and monitoring activities at the Data Space Governance level act to implement the decisions made at this level in the formation stage. For example, *GGF-3.2.02/DS/O+M: Policy Management* implements the access and use conditions that are defined in *GGF-3.1.06/DS/F: Creating a Framework for Declaring Access, Usage Policies*.

GGF- 3.2.02/DS/O+M: Policy Management: Enforcing agreed/negotiated access and use policies set by data holders in the context of requests/bids for data. ● Technical aspects: XACML or similar processes.
GGF- 3.2.03/DS/O+M: Consent Management: Managing and securing consents of data subjects (including the subjects of non-personal data)[100] ● Technical aspects: How are consents for personal data managed? How is the interaction with consent owners (e.g., persons) managed? What standards and/or solutions are used?

[95] https://internationaldataspaces.org/wp-content/uploads/dlm_uploads/IDSA-Position-Paper-Usage-Control-in-the-IDS-V3.pdf
[96] See [33] "Contractual Principles: Content"
[97] See [33] "Contractual Principles: Content"
[98] See [33] TS.3.4 Transaction management
[99] See [33] TS.6.2 Personal data management solution
[100] See [33] BO.5.3 Consent management ; TS.3.3 Data usage control solution

● Governance Aspects: How is the performance and compliance of the consent management process monitored and reported?
GGF- 3.2.04/DS/O+M: Provenance & Traceability: tracking sources, processing, dependencies ● Technical aspects: ISO 19115 provides a promising model for this (adapted from geospatial applications).
GGF- 3.2.05/DS/O+M: Orders/Requests for Data from "Data as a Service".
GGF- 3.2.06/DS/O+M: Operate Data Transfer Services: if this service is provided by the data space.
GGF- 3.2.07/DS/O+M: Transaction Logging and Usage Accounting. Logging data transactions in a secure but auditable manner, performing analysis to enable assessment of compliance and measurement of performance.
GGF- 3.2.08/DS/O+M: Composition: Enabling workflows with multiple services and pipelines
GGF- 3.2.09/DS/O+M: Infrastructure Supports: orchestration of compute & storage. Ensuring orchestration services operate as needed by users, providing helpdesk support.

4.4.Generic Data Governance

This section explores the governance framework needed to respect requirements set by data holders on the acceptable visibility, access and use of the data they hold. To a large extent, this framework reflects the typology of data that has developed as new legislation has been adopted flowing from the European Strategy for Data –which was outlined in section 2.1, specifically Data Governance Requirement *GGF-1.01/DG/F*.

4.4.1. Generic Data Governance: Formation

At the formation stage, this framework establishes a taxonomy of data types, along with a lifecycle model that structures the relevant legal and business requirements.

GGF- 4.1.01/DG/F: Define Data Typology.

- **Technical aspects**: Potentially build on W3C: Data Privacy Vocabulary
  - The W3C DPV currently addresses personal data in detail.
  - Additional data types defined by EU legislation should be added.
- Identify possible **Intellectual Property Rights** that can be associated with a given item of data:
  - Trade Secret
  - Copyright
  - Sui generis Database
  - other
- **Governance aspects**
  - Data Providers must be required to use the Data Privacy Vocabulary to accurately label the data being provided, e.g. as "personal data", and to fully disclose related information so that relevant legislation can be followed.[101]
  - If data providers wish to provide anonymized personal data they should be held responsible for the data's non-personal nature in the event of any breach.
  - Failure to fully annotate data with the DPV terms introduces the risk of non-compliance and data breaches through use of inadequate protection and confidentiality measures.
  - Constitutive Agreements should emphasise the severity of this omission with regular monitoring, enforcement and imposition of penalties and corrective action.
  - Constitutive Agreements should also specify corresponding roles and responsibilities, e.g. data controller and/or joint controllers, with clear definition of performance required, and strict monitoring (e.g. through registry of machine-readable data processing agreements that can be assessed against standards and best practices) and enforcement.[102]

GGF- 4.1.02/DG/F: Define Metadata/Self-Descriptions required for all data.

- The Data Act[103] requires that data providers must specify the "data content, […] data collection methodology, data quality and uncertainty […] sufficiently […] in a machine readable format" to allow a prospective data consumer to find, access and use the data.
- Metadata and self-descriptions should align with the data models and interoperability standards selected for the data space. (See ***GGF- 3.1.01/DS/F***.)

GGF- 4.1.03/DG/F: Define Measures of Quality and Fitness for Purpose. How to ensure that the data quality is at a sufficient level?[104] The Data Act[105] requires that data providers must specify the "data collection methodology, data quality and uncertainty […] sufficiently […] in a machine-readable format" to allow a prospective data consumer to find, access and use the data.

- Technical aspects: Characterising issues of poor quality: missing data, outdated data, metadata errors, semantic differences, real-time/latency requirements). Identifying possible corrective mechanisms. Defining responsibility for these operations.

[101] See [33] TS.6.1 Inclusion of personal data (SEC)
[102] See [33] TS.6.3 Personal data related obligations (SEC)
[103] Data Act Article 28
[104] See [33] BO.5.5 Data quality; TS.4.4 Data quality (technical implementation)
[105] Data Act Article 28

● Governance aspects: Setting standards/tiers for data providers to commit to, defining measurement and possible enforcement processes.
GGF- 4.1.03/DG/F: Define Measures of Quality and Fitness for Purpose. How to ensure that the data quality is at a sufficient level?[106] The Data Act[107] requires that data providers must specify the "data collection methodology, data quality and uncertainty [...] sufficiently [...] in a machine-readable format" to allow a prospective data consumer to find, access and use the data. ● Technical aspects: Characterising issues of poor quality: missing data, outdated data, metadata errors, semantic differences, real-time/latency requirements). Identifying possible corrective mechanisms. Defining responsibility for these operations. ● Governance aspects: Setting standards/tiers for data providers to commit to, defining measurement and possible enforcement processes.
GGF-4.1.04/DG/F: Security. What security requirements are required for an item of data: at rest, in transit, on the part of any data recipient? This may include establishing limits on where data may be moved to (e.g. only certain countries, only to storage or processing facilities with a defined level of security, etc.)
GGF-4.1.05/DG/F: Privacy/Confidentiality. What privacy and confidentiality commitments are required from processing facilities, networks, data recipients?
GGF-4.1.06/DG/F: Visibility. To whom can the metadata/self-descriptions about an item of data be exposed? Open/Public, limited to registered participants in the data space, limited to smaller groups of participants (how are these defined), only exposed to clearly identified participants authorised by the data holder. (Note: limitations on visibility and findability may be required for certain kinds of personal data.)
GGF-4.1.07/DG/F: Findability. Can the data be searched for? Or does it require the searcher to know where to look, or to ask the data holder? (Note: limitations on visibility and findability may be required for certain kinds of personal data.)

5. GDDS Governance Framework Requirements

During Phase 1 of the GREAT project, requirements from the Community of Practice gathered from the reference use cases and data sharing initiatives are highlighted where evaluation and analysis have brought up insights into the design. Reference use cases and data sharing initiatives consulted in the first phase reflect well-established data initiatives and help to understand the current landscape. Phase 2 of the project will continue with the evaluation of the GDDS requirements based on a broad set of cross-disciplinary and innovative use cases in order to bring those requirements that go beyond the state of the art. The requirements follow the structure of the Generic Governance Framework presented in Chapter 4.

5.1.GDDS Context

Chapter 2 explores the context of Data Spaces generically: how are they defined, what are they supposed to do, how do they relate to existing data, services and even data sharing initiatives? This section focuses on the Green Deal as the specific context of the GDDS.

[106] See [33] BO.5.5 Data quality; TS.4.4 Data quality (technical implementation)
[107] Data Act Article 28

The European Commission's JRC analysed the legal foundation of the Green Deal Data Space in [53]. The JRC identified six potential policy pathways, each incorporating a growing number of categories of data (see Section 2.2 and Table 1) as well as relevant horizontal EU legislation. Figure 4 illustrates these pathways.

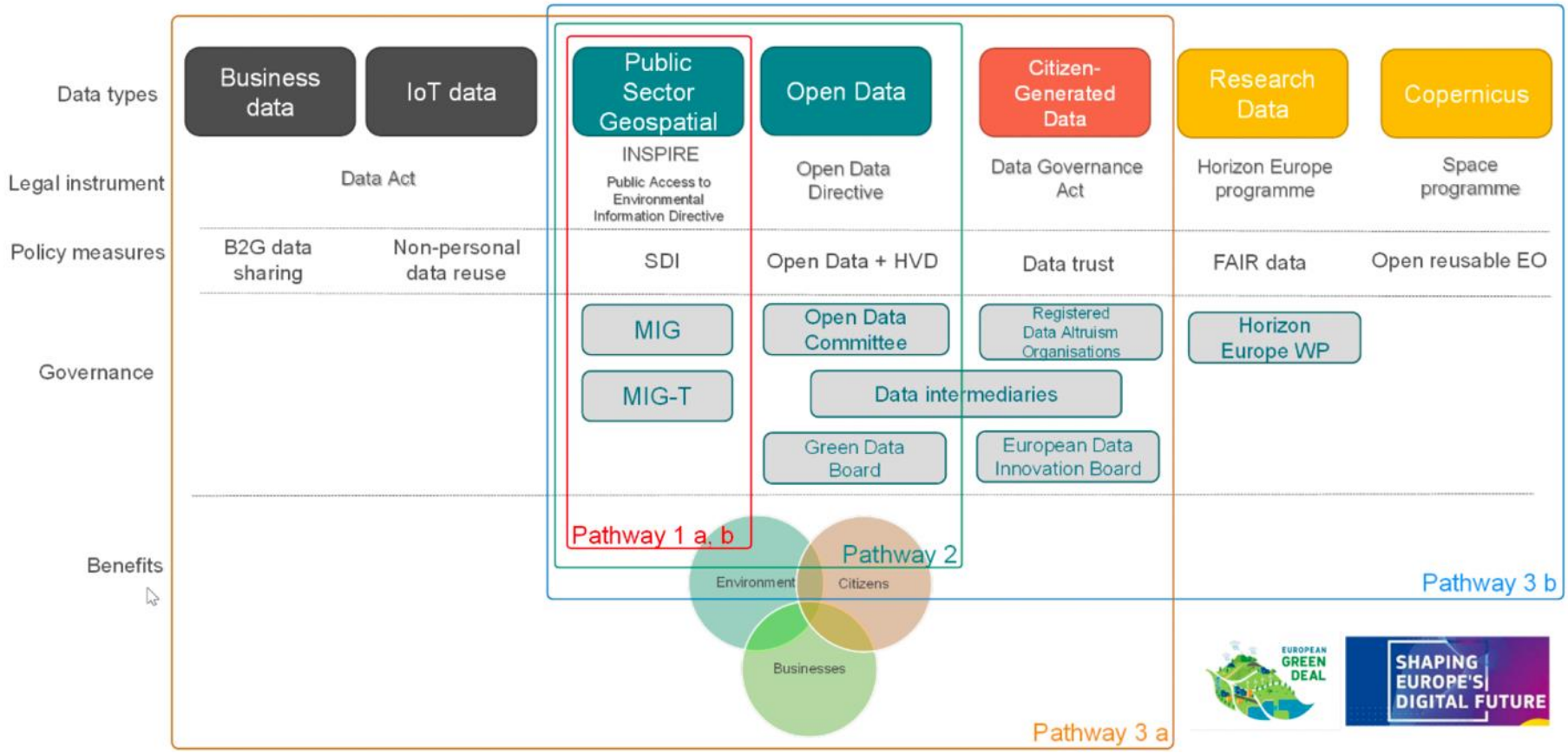

Figure 4: Policy pathways for establishment of the European Green Deal Data Space.

Pathway 4 is not pictured (it is the union of Pathways 3a and 3b) but is the most ambitious, integrating all five of the previously presented options and achieving the EC's objective of seamless data sharing regardless of data source, while respecting the principle of data sovereignty. The JRC concludes that "this pathway might turn out to be overly complex and too ambitious." However, addressing the challenges addressed by the European Green Deal will benefit from harnessing data from across all the data categories depicted, so Pathway 4 should be embedded in the long-term vision for the GDDS, despite its complexity. An ambitious, fit-for-purpose Green Deal Data Space will enable, enhance and catalyse the effective and efficient achievement of (and monitoring of progress towards) the European Green Deal.

The Green Deal Digital Ecosystem has several important characteristics that may distinguish it from other digital ecosystems:

- It is diverse. Members of the community of practice range from individuals (citizens) to profit-making and non-profit organisations, to governments and non-governmental organisations, and many more.
- There are numerous existing data providers and some existing data management services and sharing initiatives in this domain, with varying mandates and funding.
- Its scope ranges from local to global. The problems addressed by the European Green Deal are not limited to Europe – Europe must work with other regions, nations, and peoples around the world, understanding the effect of external events and aligning action at the appropriate level, with a global perspective.
- It is multi-sectoral. The problems addressed by the European Green Deal touch every sector of the economy and human activity – sometimes with opposing effects that require trade-offs.
- Within Europe, the Green Deal policy objectives to be supported by the Digital Ecosystem are very broad, as is the related regulatory framework.
- Data provides critical evidence, supporting arguments for real action with real consequences for people, so the data must be high quality, its analysis must be reliable and trustworthy, and the results must be reproducible and verifiable.
- The focus is not just on "pooling data", but on improving human well-being and the well-being of the planet in the face of imminent visible and invisible threats, requiring actionable insights, measurable results and clear accountability, against the policy objectives set forth by the European Green Deal, as well as by other international frameworks endorsed by the EU, such as the SDGs of the UN Agenda 2030.
- There will be (and are) healthy debates about the right actions to take, so respect for all participants is critical.

This diversity of stakeholders and requirements makes the planning and design of a Green Deal Data Space challenging. Broad participation in initial consultation and eventual governance is important to establish the data space as a trustworthy and legitimate initiative on which key actors will rely. At the same time, specific use cases are more likely to reach their objectives, and key

actors will be more likely to participate, if initiatives have a more focused scope and narrowed participation. Balancing these two across a single, all-purpose data space will be difficult[108].

Building on the points made in Chapter 2, the GREAT project proposes to create a set of guiding principles and governance framework for the Green Deal Digital Ecosystem overall. Within this Ecosystem, separate data space initiatives (not a single data space) that engage the necessary actors from the broader Ecosystem, both as participants and as members of the governance process. While separate, these data space initiatives would be required to maintain compatible governance structures and technical approaches. (Data Space Initiatives are defined by the DSSC as "A collaborative project of a consortium or network of committed partners to deploy and maintain a data space".)

For example, separate data space initiatives might be established to tackle marine-related challenges in the Baltic Sea, Mediterranean Sea, and the North Sea and English Channel (see Figure 5). Some data would be common across all three data spaces, for example provided by the European Marine & Data Network (EMODnet, a long-term EU data service and initiative)[109], complemented by data from local authorities, as well as locally active organisations such as offshore wind turbine operators, fishing companies, transport operators, etc. The same analytical tools and measurement frameworks could be employed in all three spaces to foster collaboration and alignment on results. Similar governance structures would be put in place for each data space initiative, but each initiative would involve different sets of actors in these governance structures. Nested governance structures could ensure efficiency while maintaining trust. This approach avoids data silos by ensuring initial compatibility of governance and technology among the different data space initiatives. Trust can be developed within the individual initiatives, enabling harmonisation and possible merger of the different initiatives into one or a few data spaces.

[108] Fritzenkötter, J., Hohoff, L., Pierri, P., Verhulst, S.G., Young, A., and Zacharzewski, A., 'Governing the Environment-Related Data Space'. *TheGovLab*, 2022, https://files. thegovlab.org/erdgovernance.pdf. Fritzenkotter et al. highlight the same diversity of interests and stakeholders, as well as the challenge of finding a productive balance among them.
[109] https://emodnet.ec.europa.eu/en

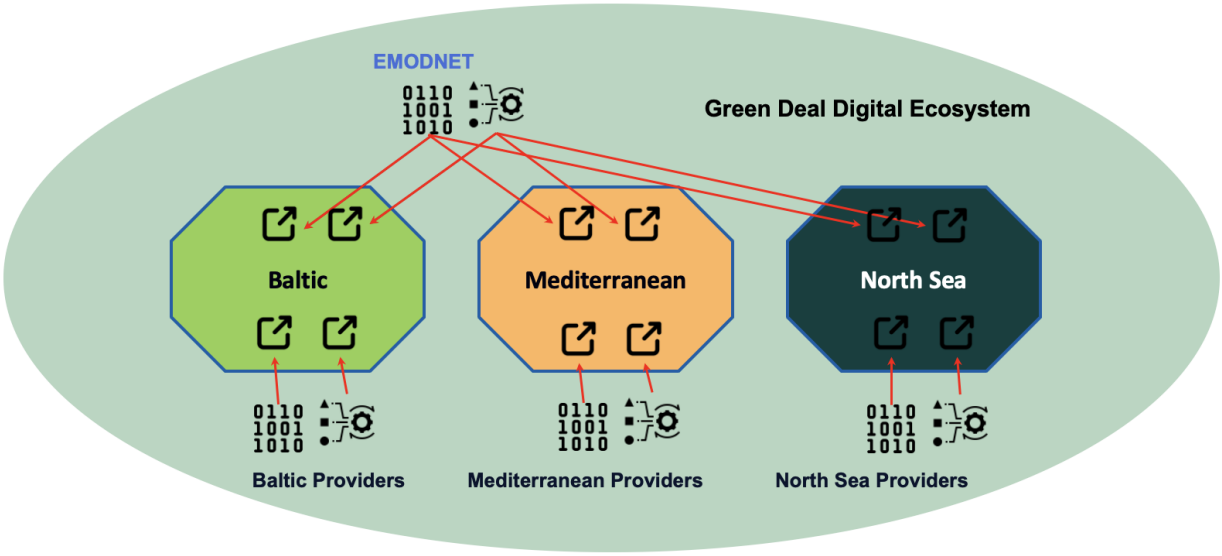

Figure 5: Multiple Data Space Initiatives within a Single Green Deal Digital Ecosystem

This approach mirrors the progressive alignment achieved by research data infrastructures in many scientific disciplines, which has in some cases required several decades of effort. Making data interoperable can take time – usually requiring new coding and metadata to be applied, as do modifications to existing technical infrastructure to enable technical interoperation. However, even when the data involved is open and FAIR, this alignment can still take time because expanding the "circle of trust" from a single discipline to a broader community is an incremental human process that requires time[110]

5.2.Mission, Objectives and Vision for the Green Deal Digital Ecosystem

Given the context described above, the GREAT project frames its consideration of mission, objectives, and vision first at the level of the Green Deal Digital Ecosystem and then as a guideline for the design and governance of each Green Deal Data Space Initiative established within that Ecosystem. Table 3 presents some preliminary approaches to these issues, which will be further refined in consultation with stakeholders in Phase 2 of the GREAT project:

Table 3: Mission, Objectives and Vision

Scope	Green Deal Digital Ecosystem (GDDE)	Green Deal Data Space Initiatives (GDDSIs)
Mission: This is the near term target – the minimum viable solution.	● Level 1 by 2026 ● GDDE Community of Practice is clearly identified and organized, with Participants that have a good understanding of their role	● Data Space Initiatives are identified, aligned with EGD strategic actions, UN SDGs. ● standard governance structure developed for the GDDE has been aligned to the needs of each data space initiative, and

[110] Interviews by the authors with representatives of three significant research data infrastructures, SeaDataNet https://www.seadatanet.org, IS-ENES https://is.enes.org/index.html, GBIF https://www.gbif.org.

	and commitment towards the Ecosystem. ● A registry of Green Deal Data Space Initiatives is available, including targeted use cases, EGD strategic actions, UN SDGs. ● Roadmaps/how to's for data space implementation are in place and tested (core services and providers identified, funding models, operations plan. ● Level 1 Technology Frameworks are defined, with compliant technologies assessed and identified, available for interoperable implementation by service providers. ● Trust framework, including trust anchors and credential providers, is in place for the full Ecosystem. ● "Easy entry" tools/processes to incorporate both public and non-public data, including data preparation and annotation tools. Tools make it easy for data providers to correctly annotate data to define access and use policies. ● Controlled mechanisms available for data providers to join multiple data space initiatives easily and confidently. ● A Standard Governance Structure is available as a template and ready for adaptation and implementation by specific segments of the Community.	agreements, governing bodies, legal entities as needed, are in place and operational. ● Operators, enablers, organizing entities, data intermediaries, etc. are in place and working using the Level 1 Technology Framework and compliant tools. ● Key data holders/providers have prepared their data for participation in identified Data Spaces.
Objectives: What the Green Deal Digital	● GD Digital Ecosystem Community will develop the transformation tools, analytics, etc. to enable	● Each Data Space Initiative will increase its objective level by

Ecosystem will achieve over time	increased objective levels as follows: ○ Level 3 by 2028 ○ Level 5 by 2030 ○ Level 7 by 2032	adopting/developing new tools and value-added functions. ● Each Data Space Initiative will be encouraged to achieve increased integration with other Data Space Initiatives and to begin to form integrated Data Spaces. ● Sister data spaces will engage in cross-fertilization, harmonisation, and possible merge/consolidation.
Vision:	● GD Community is strong and self-sustaining. ● GD Community spans the global north and south. ● GDDE insights are driving positive action around the world. ● GDDE and GDDSIs have synergistic business models that are self-sustaining through capture of value generated.	● Individual data spaces have consolidated into relatively few. ● Links with data spaces in other sectors are in place and busy, with data being provided and consumed in both directions.

5.3. GDDS Legal and Regulatory Context: Sector-Specific Legislation and Regulation

The INSPIRE Directive clearly represents legislation that is particularly relevant for the Green Deal.

GDDS-1.09/DS/F: The GDDS supports INSPIRE/HVD compliant data/services provided by public administrations.

EMODnet: Contributors to the EMODnet data portal provide INSPIRE compliant services, e.g., via OGC data access services and metadata descriptions. Access to EMODnet from the GDDS can be provided via existing standardised INSPIRE compliant interfaces.

5.4.GDDS Digital Platform Governance

5.4.1. GDDS Digital Platform Formation

5.4.1.1. GDDS Digital Platform: Landscape Design

GDDS-2.1.01/DP/F: Landscape Design: The GDDS shall integrate existing data infrastructures from thematic CoP and data initiatives with regional, national, European and ultimately global scopes

The landscape of GDDS stakeholders and existing data initiatives ranges in diversity and maturity levels. The GDDS shall support initiatives based on infrastructures that may differ on objectives, resources, and governance frameworks provided that these do not conflict with the GDDS overarching framework.

As of Phase 1, GREAT has done a deep analysis of five reference use cases and a number of data sharing initiatives, consulted with 8 data initiatives, 10 data spaces and 57 stakeholders and 7 Horizon Europe projects to understand the current data landscape in Europe and beyond

5.4.1.2. GDDS Digital Platform: Mission and Objectives

GDDS-2.2.01/DP/F: Objectives: The GDDS shall follow the Data Spaces Objectives taxonomy as defined in Table 2 accommodating the diversity of stakeholders, scale up requirements and incremental implementation.

The following high-level requirements have been identified by the reference use cases and data-sharing initiatives and mapped to the Objective Taxonomy Levels. These requirements represent the problems and needs that consulted stakeholders will look towards the GDDS to solve.

Table 4: Requirements from User Cases Mapped to Objective Levels

Stakeholder	Requirements
EPOS	● Enabling data sovereignty services e.g., Authentication and Authorisation Infrastructure (L1) ● Enabling multi-stakeholder reward mechanisms (profit/citations etc.) (L2) ● Focus on potential users, scientists at the forefront of use cases to solve scientific challenges (L4) ● Support the application of AI to data (L4) ● Support with exploitation of community solutions (e.g., EPOS portal) (L4)
EMODnet	● Enable M2M interoperability with other sectors (L2) ● To have EMODnet's profile raised as the marine in situ data service of the European Commission. (L0) ● To have the GDDS accept EMODnet data publishing technologies, which are based on recognized standards.(L2)

GOS4M	● Increase and support the role of EO data in the evaluation of policy frameworks (L5) ● Provide trusted solutions based on digital infrastructures and AI to facilitate evidence-based decisions (L5) ● Provide access to data from other sectoral domains e.g., health data space (L1)
WATER TF	● Data consumers have easily accessible and discoverable data to input for hydrological models (L1) ● Data providers get enhanced visibility and ensure access and discoverability of model outputs (L0 & L1)
BIOGIS	● Data consumers have easily accessible biodiversity data from global, local sources including citizen science and satellite images (L1) ● Enable access to and use of biodiversity data for decision making e.g., supporting industrial set up of solar and wind energy plants (L5)

GDDS-2.2.02/DP/F: Use Cases: The GDDS shall accommodate a range of diverse use cases, covering thematic or domain specific, cross-disciplinary, cross-border, multi stakeholders and innovative aspects, as well as leverage existing data management and sharing services and initiatives.

During Phase 1 of the project, GREAT selected five reference use cases and initiatives and set up related Task Forces to consult with relevant communities across different themes, domains and stakeholder types. (See Figure 5.) These use cases and initiatives were selected according to:

1. Relevance to the Green Deal,
2. Diversity of strategic actions,
3. Relevance to the GD policies and legislations,
4. Readiness of the solution,
5. Involved actors,
6. Geographical scope.

During the consultations, different stakeholders were interviewed on different topics, e.g., including technical and governance roles.

Table 5: Reference Use Cases Considered by GREAT in Phase 1

Use case/ initiative/ Task Force	Theme	Stakeholders	Scope	Strategic Action	Readiness
EPOS	Solid Earth	Research Infrastructures	European	-Zero pollution -Climate Change adaptation	Operational

EMODnet	Marine (in situ)	EMODnet Secretariat, (commissioned by the European Commission)	European	- Biodiversity - Zero pollution - Climate change adaptation	Operational
GOS4M	Pollution (Mercury)	Global Contributors from Mercury monitoring network	Global	-Zero pollution	Operational
WATER TF	Hydrology	- Researchers: Utrecht University (UU) and Helmholtz Centre for Environmental Research (UFZ) - International Organisations (ECMWF) - United Nations Bodies (WMO) - European Agencies (JRC)	-WMO Hydrology (Global) -UU (Global Hydrology model) -JRC (European Hydrology model) -ECMWF (European Hydrology model)	- Zero pollution - Climate change adaptation	Development
BIOGIS	Bio-diversity	Industry	National	- Biodiversity	Development
Marine Task Force	Marine	- EMODnet network - Copernicus Marine - SeaDataNet - Blue-Cloud - IMEC	European	- Biodiversity - Zero pollution - Climate change adaptation	Operational

Principles were formulated with the collective input from the GREAT Consortium in a workshop in Utrecht (March 2023)

GDDS-2.2.03/DP/F: Principles, Values: The GDDS adopts the following principles and values:

- GDDS enables seamless data sharing by stakeholders who wish to contribute to the European Green Deal objectives.
- GDDS builds on existing data sharing initiatives from all scopes (local, regional, national, European, global) where communities are already established.
- GDDS enables FAIR data principles.
- GDDS ensures data sovereignty in all data sharing transactions.

- GDDS offers transparent, open and clear rules for the governance of the data space and ethical usage of the data.
- GDDS adheres to European values including security, privacy preservation, trust and fair competition.
- GDDS offers a flexible design capable of adapting to the technical and governance evolution.
- GDDS promotes collaboration, digital, scientific, and entrepreneurial innovation.
- GDDS abides by the EU Do No Significant Harm principle according to the EU taxonomy for sustainable activities.

GDDS-2.2.04/DP/F: Actors Needed, Desired: The GDDS must engage with stakeholders from multiple sectors and across all levels. These include representation at the governance and technical level, e.g. service/data providers and consumers, infrastructure providers and operators, etc. Stakeholders should include representatives from the quadruple helix framework[111] that is public administrations, research infrastructures, industry, and citizens and should be verified as trustable entities before they contribute to the GDDS

Reference use cases and initiatives have been analysed considering the diversity of stakeholder types (see Table 6 below).

GDDS-2.2.05&06/DP/F: Roles Needed, Desired & Role definitions, expectations: At the Digital Platform level the GDDS aligns with the following roles, defined by the DSSC:

- **Data space enabler**: A data space participant that provides a (technical or non-technical) service enabling data transactions for the transaction participants while not directly participating in that transaction itself. Examples of enabling services include identity provisioning, vocabulary provisioning, interconnecting, clearing, etc.
- **Data rights holder:** A transaction participant that has the legal right to use, grant access to or share certain data.
- **Data provider:** A transaction participant that, in the context of a specific data transaction, technically provides data to the data receivers that have a right or duty (granted by the data rights holder) to access and/or receive that data.
- **Data receiver**: A transaction participant to whom data is, or is to be technically supplied by a data provider in the context of a specific data transaction.
- **Data user**: A transaction participant that has been granted (lawful) access and the right to use data as the result of a specific data transaction. Also known as data rights receiver.
- **Data intermediary**: A data space enabler that (technically and legally) connects one or more data space participants to the data space, thereby enabling them to establish relationships and execute data transactions with other participants in the data space.

Dedicated roles from the reference use cases and initiatives have been identified and defined (definitions provided for EPOS, EMODnet and GOS4M) in section 4. A mapping is presented in Table 6 where the corresponding role exists. Dedicated governance roles have not been defined at the digital platform level. These need to be addressed according to the governing bodies in place.

[111] https://en.wikipedia.org/wiki/Quadruple_and_quintuple_innovation_helix_framework

Table 6: GDDS Use Case Roles Mapped to DSSC Roles

Use case	ROLES
EPOS	• Thematic Core Services -> Data providers • Data curators -> Data Space enablers • System administrators -> Data Space enablers • End user -> Data user • IS-Central Hub -> Data intermediary, Governance role
EMODnet	• EMODnet Secretariat, Thematic & Data Ingestion partners -> Data space enabler, Governance Role • EMODnet data providers ->Data right holders • European Commission > Data right holder • EMODnet Steering Committee-> Governance role • EMODnet Technical Working Groups -> Governance role • EMODnet Marine Knowledge Expert Groups -> Governance role • EMODnet users: Data user
GOS4M	• Steering Committee -> Governance Role • Scientific Advisory Board -> Governance role • Expert Groups – Governance role • Regional GEO -> Data providers • GOS4M Knowledge Hub - > Data intermediary
WATER TF	• Meteorological Data Ingestion -> Data users • Hydrological modeller -> Data user, Data provider • Water manager and policy maker -> Data user
BIOGIS	• Biodiversity Data ingestion -> Data user • BioGIS web viewer -> Data Space enabler (service provider)

GDDS-2.2.07/DP/F: Resources Needed, Desired: GDDS must accommodate data sources as identified in the High priority datasets inventory (D5.1) and provide the infrastructure needed to deploy the Blueprint Reference architecture D3.1 and minimum viable Data Space.
GDDS-2.2.08/DP/F: Service "composition" and "stacking" required: GDDS shall provide information of interoperable data/services to enable workflows, dataflows access and orchestration e.g., geospatial datasets should state interoperability with INSPIRE compliant services

5.4.1.3. GDDS Digital Platform: Value Creation

GDDS- 2.3.01/DP/F: Targeting, Awareness, Inclusion & Retention: The GDDS shall establish the Green Deal Community of Practice with the intent to identify representative problems and needs and collaboratively build solutions to those community needs. A stakeholder needs focus perspective should ensure active participation, inclusion and retention in the ecosystem.

GDDS-2.3.02/DP/F: Value sharing, exploitation, collective action, generativity: In alignment with the DSSC, GDDS shall enable value creation at the individual business model level, collaborative and infrastructure level.

Collaborative business models include primarily the fulfilment of a common societal good, in this case as a consequence of the fulfilment of the Green Deal policies e.g. clean air, soil and water, clean energy and mobility, sustainable agriculture, supplier chains, cities and buildings and an equitable sustainable society in general. In addition, GDDS collaborative business models will enable joint innovation, collaborations, scientific excellence, cost of ownership savings due to the sharing of components, further outreach via connectivity, promotion and scale up of data/services beyond the usual scope of operation.

Individual business models shall be enabled by the individual use cases and applications across the whole Green Deal spectrum. The value proposition will vary depending on the needs from the actors involved. During consultations with stakeholders, research communities highlighted the value of rewards mechanisms such as citations and promotion of Open Data and Open Science while industry reflected on the legal obligation of generating profits.

GDDS should assess individual business models to ensure that the intended purpose respects ethical considerations for data usage.

Table 7 reflects on the value proposition from each of the individual reference use cases and initiatives.

Table 7: Value Propositions by Use Case

Use case	Value proposition, business goals, cost models
EPOS	• Enabling efficient and unified access to solid Earth science data from different sources across Europe. • Provide easy data discovery, access, use and reuse for researchers, educators, and policy makers in the field of solid Earth science. • Cost model: Open access to data • Funding: Membership Fees (Members and Observer) and hosts contributions for the operation of Executive Coordination Office (ECO) and Integrated Core Services Central Hub (ICS-C)
EMODnet	• The **European Marine Observation and Data network** (**EMODnet**) is a **partnership** of more than **120 organisations** which assemble and disseminate marine in-situ (field) data, data products and services in **7 thematic disciplines** (bathymetry, biology, geology, chemistry, physics, seabed habitats, human activities) using **FAIR** (***f****indable*, ***a****ccessible*, ***i****nteroperable*, ***r****eusable*) **principles**. Across the 7 disciplines, **EMODnet** makes available quality-assessed data and data products of hundreds of parameters. The partnership includes neighbouring countries as well as those in the EU so as to deliver complete coverage of all European seas. A secretariat service is responsible for the overall management and communication, and an ingestion service assists data holders with standardising their data and opening them for reuse. The original datasets openly available through **EMODnet** are *intellectual property of the organisations that originally collected them or commissioned their collection*. **EMODnet** establishes agreements with these organisations (through the procurement process and beyond), which accept to make their data openly available for use, usually under no or minimum restrictions. **EMODnet partnership** creates consistent metadata and uses these datasets to create **harmonised data products** in the level of the European sea basins, *which are intellectual property of the* ***European Commission*** (under **Creative Commons** licence). The **European Commission** selects the partnerships through open calls for tender. • Cost model: Free and unrestricted access to data. • Funding: EC service (DG-MARE) funded by the European Maritime, Fisheries and Aquaculture Fund (EMFAF[112]).
GOS4M	• To promote the establishment of a federation of existing regional and global monitoring networks that would allow to provide global comparable monitoring data for the purpose of the Minamata Convention on Mercury (MCM) • Cost model: Open access • Funding: Flagship participants

[112] https://oceans-and-fisheries.ec.europa.eu/funding/emfaf_en

WATER TF	• Provide a global estimate of global water availability for policy support with accurate estimates of the amount of water stored underground, in lakes, rivers and streams using state-of-the-art hydrological models. This allows to better make policy decisions based on these estimations when observations are absent (Global hydrology) • Provide seasonal forecast information to make more accurate decisions in water management and drought mitigation, specifically aimed at water managers. Make predictions for the coming weeks to months (seasonal forecasting) • Cost model: Open data • Funding: Research projects
BIOGIS	• Allow large companies and multinationals to monitor and map the pre and post biodiversity level and the relative impact of their interventions on the territory (also with the support of satellite data) • Provide a global certified geo database with all the data needed • Cost model: Freemium • Funding: For profit

GDDS-2.3.03/DP/F: User Experience: GDDS shall ensure optimal user experience at end user interfaces including graphical interfaces and M2M APIs in order to fulfil its objectives

During consultations at the stakeholder forum, the majority of participants from the Community of Practice agreed that ease of use and interoperability of M2M APIs across systems would be very important.

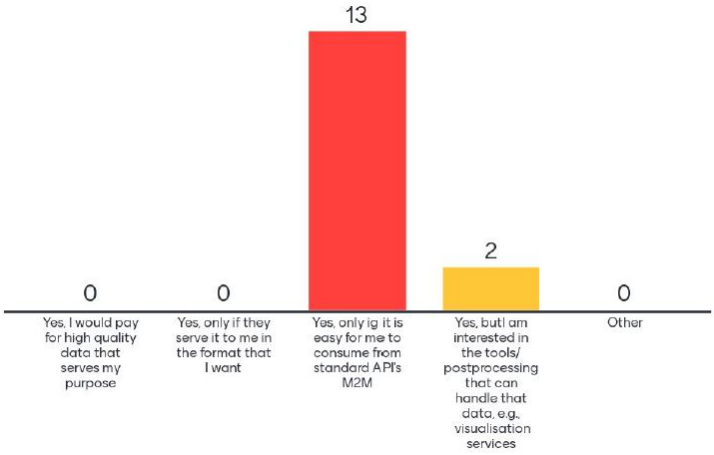

Figure 6: Results of the stakeholders consultations

5.4.1.4. GDDS Digital Platform: Technical Architecture and Control

GDDS- 2.4.01/DP/F: Architecture, Control: The GDDS adopts the digital ecosystems architectural approach taking into account the dynamic nature of evolving ecosystems and the autonomy of contributing systems, bringing together an overarching system that will enable multiple interdisciplinary use cases

Design principles of the ecosystem as well as the architectural design of the technical blueprint have been defined in GREAT D3.1 "Initial Blueprint of the GDDS Reference Architecture". These include inclusiveness, fairness and autonomy at the ecosystem level and Lower Entry barrier, System of systems approach, Standardization and Mediation, Data as Entry Point, Loose Coupling and Interoperability/Security Orthogonality at the technical design level.

GDDS- 2.4.02/DP/F: Openness, Modularity, Intellectual Property Policy: The GDDS shall promote practices to maximise uptake and minimise barriers to adoption, adaptation and reuse. Data space technical components shall be provided by Free and Open-Source Software licences.

Dedicated arrangements may exist where IPR or personal data require specific processing e.g. data curation, anonymisation, pseudonymisation, etc.
Integration of IP from service providers into marketplaces or intermediaries should be taken into account and managed in a dedicated process to ensure a coordinated and efficient multi-party IP software and services operation.

GDDS- 2.4.03/DP/F: Centralised, Federated, Distributed: GDDS infrastructure shall provide and federate access to multiple distributed data sources. Dedicated components of the architecture may follow a centralised or distributed deployment, e.g. catalogues , marketplace or harmonisation services.
GDDS- 2.4.04/DP/F: Participant/User Identification, Registration, Trust Framework: GDDS shall provide federated access control to distributed data sources via means of well identified and trusted participants, with the necessary access and usage control mechanisms required by each data provider.
The GDDS should ensure that data providers also comply with the trust framework and that their offerings align with overall value proposition and objectives.
A minimum set of rules to ensure trust and compliance to the **GDDS acceptance criteria** should include the following:

- Data providers shall represent either a natural person, with verifiable credentials, a legal entity, an EU represented institution, project or an accredited international collaboration with representation in Europe.
- Data providers shall disclose Terms and Conditions/Licensing terms of access and usage to their data/services.
- Data providers shall disclose the level of data quality and quality assurance processes their data and services have been subject to.
- Data providers shall comply with legal and ethical requirements ensuring that they have the legal right to share the data and that they do not infringe any copyrights, patents, or other intellectual property rights. Transparent and ethical data handling are essential.

- Data providers should provide data that is relevant to the fulfilment of the Green Deal objectives.
- Data providers shall provide access to data via machine readable APIs.
- Data providers shall make available metadata about their data/services to ensure semantic understanding of their offerings.
- Data providers should implement appropriate security measures to protect the data from unauthorised access, tampering or breaches while in transit and storage.
- Data providers infrastructure and systems should be capable of handling the expected data volumes and provide data in a timely manner.

GDDS- 2.4.05/DP/F: Cybersecurity: GDDS shall require confidentiality, integrity, traceability and availability of the services and data made available via the infrastructure. GDDS shall comply with the NIS directive in its role as digital services operator and require that any providers of critical infrastructures available through the GDDS shall comply with applicable cybersecurity and resilience requirements. The GDDS shall ensure that security risk management, incident response, monitoring and relevant training processes are in place, with responsibilities clearly assigned and accepted by participating service providers.

GDDS- 2.4.07/DP/F: Boundary Resources: GDDS shall contemplate all resource types and its interdependencies at internal and external level in order to operate and govern the infrastructure.

Application Boundary Resources: These are requirements for i) data providers in order to be part of the infrastructure such as the technical requirements within the GDDS acceptance criteria, including provision of well defined APIs and metadata for data discovery and access and ii) for data consumers in order to get access to the data and services such as acceptance of Terms and Conditions, Licensing terms or credentials from verifiable providers, iii) for data intermediaries compliance to DGA if operating as such.

Development Boundary Resources: These resources include collaborative development environments. A DevOps methodology should ensure that developments occur in an iterative way and they can be deployed quickly to operations after testing at all levels.

Social Boundary Resources: The GGDS shall provide training, capacity building and awareness programs to ensure adoption and growth of the Community of Practice. These include training to providers that will need an understanding of the requirements needed to onboard data and services as well as consumers to identify, access and reuse data in an effective manner.

GDDS-2.4.08/DP/F: Development Plan: GDDS shall produce and maintain at all times a development and integration plan reflecting the blueprint architecture and its governance structure. The development plan should include the targets objectives, how requirements are addressed, well defined timelines and how to measure progress of implementation.

The GREAT roadmap addressed in Phase 1 (D6.1) and Phase 2 (D6.2) will present the development and deployment plan for the implementation of the GDDS.

GDDS-2.4.09/DP/F: Service Management System: GDDS digital platform shall provide a well-defined set of services to its participants managed via a service management framework. A service portfolio should be made publicly available and maintained to ensure the operational running of the GDDS as a digital platform. Well defined processes should include amongst others change/risk management, information security, service availability, risk management, reporting obligations, etc.

Key characteristics of the service offering that are a high priority to the users should be well understood. E.g., Seasonal forecasting of water resources relies on adequate data transfers and bandwidth to ensure delivery of the forecast in a timely manner. Most of the reference use cases highlighted data availability as one key offerings from data providers followed by a description of the quality of the data.

EPOS RI characterises risks into three main groups: i) Risks related to Data and Service provision, ii) Risks related to EPOS impact on science and iii) Risks related to EPOS-ERIC governance and operation. Each of the risk groups is divided into subgroups, corresponding to the qualifying dimensions which include:

1. Governance dimension
2. Financial dimension
3. Legal dimension
4. Technical dimension
5. Users dimension
6. Stakeholders (private sector, society) dimension
7. Global dimension

GDDS-2.4.10/DP/F: Business Case and Model: GDDS digital platform will initiate its deployment via Digital Europe funding program. The funding mechanisms will need to be complemented by Member States and other contributing stakeholders including research and industry.

Funding mechanisms ensuring long term sustainability of the platform are strongly related to the type of governance establishment that will be pursued. Appendix II presents an assessment of Multistakeholder Alliances and Social Enterprises as governance entities that have been used for data collaborations.

At the time of writing, several data spaces e.g., Agriculture, Language and Mobility are pursuing the establishment of a European Digital Infrastructure Consortium (EDIC). EDICS are legal instruments proposed by the Commission that will help in the implementation of multi-country projects contributing to the objectives of the Digital Decade Policy program 2030, in this case a common data space infrastructure and corresponding services. EDICS should have representation

of at least three Member States and other public or private entities can become members but cannot outvote MS.

As presented in **GDDS-2.3.03/DP/F, the** GDDS digital platform must support various types of business models and mechanisms for value generation and sharing. While the Green Deal objectives should primarily become a common societal and environmental good, investments and innovation from many stakeholders will be required to tackle climate and environmental challenges that humanity is facing.

The GDDS should pursue the establishment of a governance structure that will primarily agree to the fulfilment of the GDDS objectives while representing the interests of its stakeholders.

5.4.2. Generic Digital Platform: Governance Architecture

The following section presents a preliminary set of requirements needed for the governance architecture, including the entities that need governing and the corresponding bodies establishing the rules.

GDDS-2.5.01-06/DP/F: The GDDS will comprise an ecosystem of governed entities including:

- Community of Practice: The CoP is the main pillar of the GDDS, and its governance shall include community management, engagement, consultations, co-design, training and upskilling on a continuous cycle to ensure that the needs of all the stakeholders are always reflected in the GDDS.
- Data Providers: Governance shall include a body that ensures the onboarding of data providers complies to the defined acceptance criteria. **(GDDS- 2.4.04)**
- Digital Platform: GDDS Governance shall include the technical and scientific expert bodies to ensure that the digital platform(s) fulfils the objectives and aligns with the needs of the CoP.
- Organising entity: GDDS shall ensure that the organising entity establishes the high-level decision-making bodies such as a steering committee, board or similar, and all supervisory, compliance and operational bodies to cover finance and budget aspects, conflict resolution, ethics and privacy, risks monitoring, outreach and communication activities. The Steering Committee is responsible for the strategic direction, mission and overall vision. It is composed of representatives from key stakeholder groups including data providers, consumers, industry, MS experts and regulatory authorities.

Section 6 describes the current governance architecture for EMODnet, EPOS and GOS4M including the governing bodies and roles. Only EPOS is established as a legal entity as an ERIC.

Stakeholders (data and service providers) were asked “would they contribute with data and services to the GDDS?” Responses provided in Figure 6:

- 39% of respondents would do so if integration is not needed on their side.
- 33% of respondents would need further information to be able to decide.
- Only one respondent was willing to contribute to the strategic direction of the GDDS.

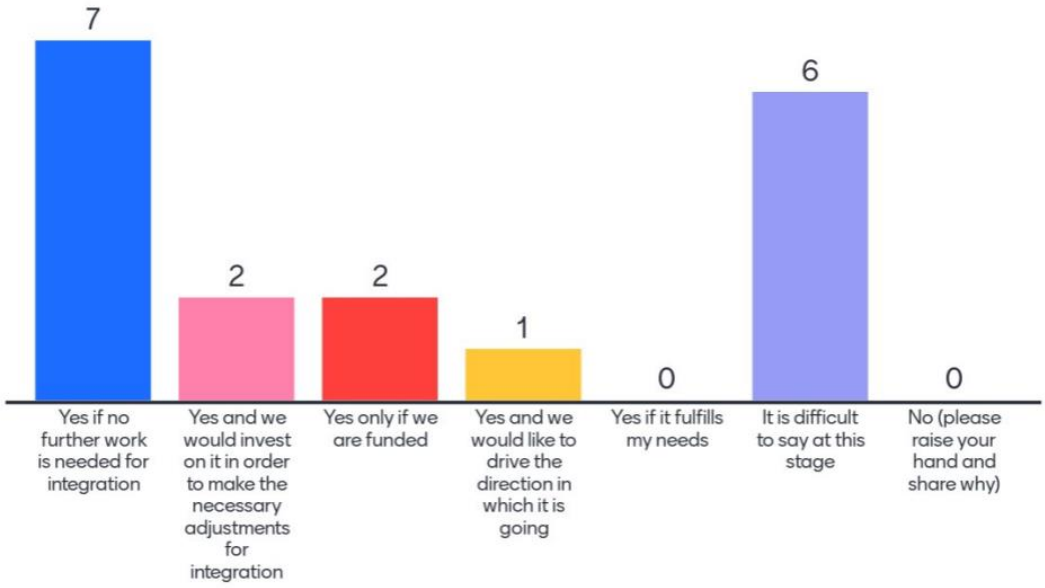

Figure 7: Results of the stakeholders consultations

Consultations highlighted: i) the importance of managing the CoP. Significant efforts should be made to bring awareness and training to the CoP in order to ensure early adoption and reach a critical mass, ii) support for legacy systems with integration work provided by the GDDS.

5.4.3. GDDS Digital Platform Formation – Launch Milestone

GDDS-2.6.01-08/DP/F-L: GDDS shall create a strategic platform launch ensuring that the governance architecture is in place and the digital platform is operations ready. The platform launch should include a minimum data and service offering to gain early user acceptance.
The launch event should address all operational, communications and support aspects including:

- Fully functional platform: Full test plan of the platform shall verify that the platform implements all functionality according to specifications and in agreement with all security measures.
- Communications, Content, Media and event launch: A launch event, press releases, web content, outreach via social media, including blogs, videos and tutorials shall be in place to ensure maximum outreach.
- Customer support and helpdesk should be in place to ensure handling of user queries and customer readiness.

5.4.4. GDDS Digital Platform Operations and Monitoring

GDDS-2.7.01-09/DP/O&M: GDDS shall monitor the effectiveness of the decision-making process in operations across the governance boards ensuring representation of stakeholders,

effective communications channels, assessing the impact of each decision and evaluating its outcomes.

Technical operational boards shall maintain and track the development plan and assess evolution in alignment with the stakeholders and CoP needs while ensuring fulfilment of the objectives.

Onboarding of participants, data and services shall comply with the acceptance criteria and automatic dashboards shall display information and metrics on the growth of the platform.

Assessment of risks and incident reports including any breaches of security measures or regulatory compliance shall be reported on a regular basis.

Operational processes defined in the service management plan shall be audited and developed accordingly to maintain compliance, continuous improvement and innovation of the services needed to operate the digital platform.

Monitoring and delivery of training activities targeting platform users, including providers, consumers or data intermediaries shall be in place.

5.5.GDDS Data Space Governance

5.5.1. GDDS Data Space Governance: Formation

GDDS-3.1.01/DS/F: **Domain Data Models, Interoperability Standards:** GDDS will support domain specific community standards where they are already established and provide means of interoperability across domains on “as needed basis” enabled by cross-disciplinary use cases.

Data providers shall specify the data/metadata formats and service interfaces at onboarding time.

GDDS will support community specific approaches to interoperability and will promote these practices across data spaces when successful. E.g. The definition of essential variables in dedicated domains, Climate, Ocean, Biodiversity provides a common understanding across data providers leading to standard encodings even if different formats (json, xml, binary) or data models are used but conveying the same semantic meaning.

D3.1 Initial Blueprint of the GDDS Reference Architecture, Annex A provides an initial list of possible service interfaces and data/metadata models to be supported in GDDS.

As described in D3.1 (Section 5.1.1) Data Transformer service providers will provide the necessary data transformations and harmonisations required by the use cases.

Figure 7 displays the data types, formats and service interfaces supported by EPOS. Harmonisation components are dealt with by the Integrated Core services Central Hub for each of the Thematic core Services Contributions.

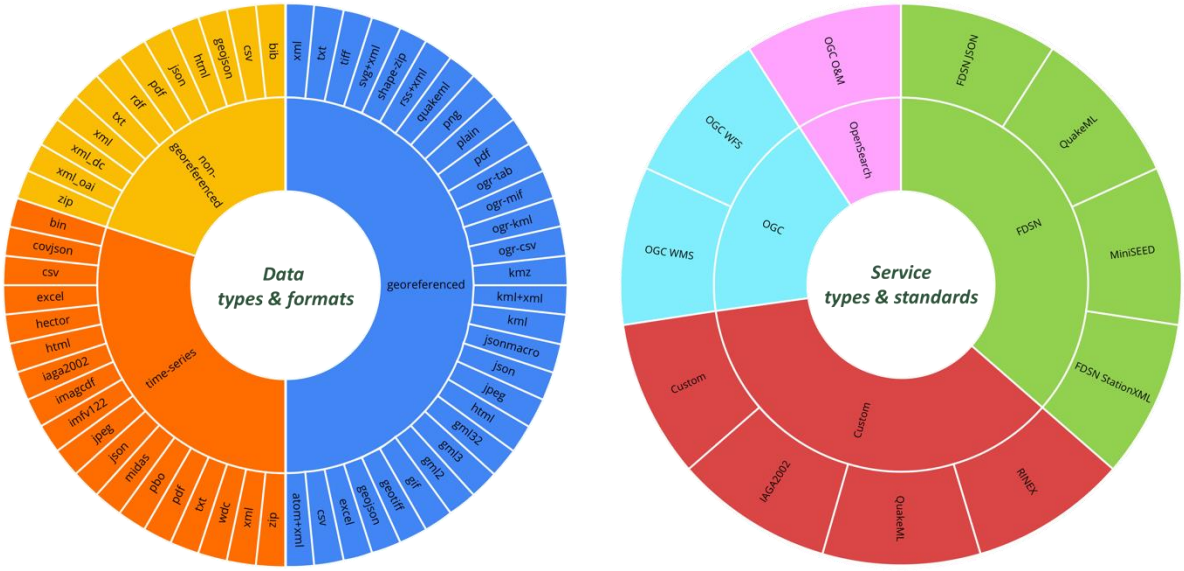

Figure 8: Data Formats and Standards supported by EPOS

GDDS-3.1.02/DS/F: Intermediation/ Marketplace/ Catalogues: GDDS shall enable data transactions via means of a Data Source Registry that enables registration of Data Sources in the GDDS and a Data Catalogue enabling discovery of Data Sources and subsequent Data Access, Transfer or Processing.

The Registry and Catalogue are both Core Components of the GDDS. (See Section 4.5 of D3.1)

Marketplaces functionality may be considered for the provision of value-added services and applications of third-party service providers.

GDDS-3.1.03-05/DS/F: GDDS should support ancillary or other services needed to enable higher levels of Purpose as defined in the Objective taxonomy. These may include anonymisation, encryption, dedicated postprocessing (sub-setting, integration, fusion, analysis, or AI/ML), compression or trend analysis services according to the specific needs of the data typology and use case definition.

These services would be part of the Facilitators component layer (See Section 4.5 of D3.1)

GDDS-3.1.06/DS/F: Defining Access, Usage Policies: GDDS shall provide federated access to distributed data sources ensuring that access and usage policies provided by data holders are respected in all data transactions.

It should be transparent for a user to discover the conditions of access and use for any data source made available to the GDDS as well as the process to obtain the necessary rights if the data is subject to any restrictions.

GDDS may establish standard terms and conditions for data transactions in human and machine-readable format if these are generalised and adopted across the data holders, providing when needed the necessary translations from legacy systems to a standardised approach.

GDDS-3.1.07/DS/F: Define the Process for Ordering Data or Requesting Data from "Data as a Service" Services: GDDS shall provide a well-defined process for service and data orders seamlessly across all the federated providers. Ordering should include open data, open access and paid for data and services.

GDDS-3.1.08/DS/F: Identify supported mechanisms for Data Transfer: GDDS shall support Data Transfer mechanisms to enable the required data transactions. These may involve transfers from original data providers to local environments or Cloud and HPC infrastructures offering computational resources or analytical environments.

The GDDS shall inform the user of the data volumes involved prior to each data transfer and when possible, an estimation of the data transfer timings based on average bandwidth availability. The data transfers may include ad hoc arrangements or regular transfers for real time data delivery with frequent updates.

The Data Mover, defined as part of the Facilitator Components in D3.1 Blueprint Architecture, will fulfil this function.

GGF-3.1.09/DS/F: Transaction Logging and Usage Accounting: GDDS shall record data transactions with logging mechanisms and provide tools or dashboards to consult the logs to monitor compliance and accounting. Data flows and usage shall be monitored, including monetised flows.

The Auditor component defined as part of Security architecture in D3.1 Blueprint Architecture will fulfil this function.

5.6.GDDS Data Governance

5.6.1. GDDS Data Governance: Formation

GDDS- 4.1.01/DG/F: Define Data Typology: GDDS shall ensure compliance to EU legislation for the corresponding data typologies E.g., GDPR compliance for personal data or Data Act compliance for industrial data, etc.

Data providers will be required to provide the typology of the data offered via the GDDS at onboarding time. GDDS shall ensure that a full taxonomy of data types is available to the data providers for classification and to the consumers to facilitate discovery and access.

GDDS- 4.1.02/DG/F: Define Metadata/Self-Descriptions required for all data: GDDS shall require metadata descriptions for all data sources onboarded to the data space. Metadata should be human and machine readable.

The GDDS shall provide metadata harmonisation processes to be able to consider well established metadata standards from different domains. A minimum core set of mandatory metadata elements necessary for discovery, access and use conditions shall be identified and provided by data providers regardless of the metadata format.

GDDS- 4.1.03/DG/F: Define Measures of Quality and Fitness for Purpose: GDDS shall require a description of the quality processes that data has been subject to as part of the metadata descriptions. GDDS shall provide the means for filtering out data sources based on the quality assured by data providers.

GDDS shall ensure that metadata made available complies with minimum set of quality standards and is fit for purpose.

The Metadata Enhancer defined as part of the Facilitator Components in D3.1 Blueprint Architecture will fulfil this function.

For example, Europeana, the data space for cultural heritage, defines a publishing framework[113] with three metadata quality tiers covering, language, contextual classes and enabling elements to enhance the user experience with multilanguage services and the best possible findability and information retrieval.

The World Meteorological Organization, WMO, provides a set of KPIs[114] to qualify the compliance with WMO Core profile and the quality of the metadata provided by national meteorological and hydrological services worldwide to the WMO Information System (WIS).

GDDS-4.1.04-05/DG/F: Security, privacy and confidentiality: GDDS shall ensure that security requirements including privacy and confidentiality for a given data source are preserved across all the data value chain.

Data providers shall ensure that security requirements are enforced at the source of origin. If data needs to be collocated or transferred to another location, e.g, for cross-disciplinary analysis, access and usage conditions as well as all security requirements shall be enforced in the target environment and while in transit.

[113] https://pro.europeana.eu/post/developing-a-metadata-standard-for-digital-culture-the-story-of-the-europeana-publishing-framework

[114] https://github.com/wmo-im/pywcmp

GGF-4.1.06-07/DG/F: Visibility and Findability: GDDS is an open ecosystem and as such visibility and findability of data providers and metadata descriptions of data sources shall be by default open to the public even if the data referred to is subject to access restrictions.

GDDS may need special restrictions for personal or sensitive data sources if data holders require to keep visibility and findability restricted to a close group. Alternatively, data sources would need anonymisation, or pseudo-anonymisation to open visibility and findability to a wider audience.

Search functionality should accommodate the needs of domain experts and non-experts end users. Learning algorithms and recommendation engines may enhance the user experience based on users' profile and previous searching experiences if they consent to do so.

6. Existing governance models from use cases and data initiatives

The sections that follow explore governance models now in use by major stakeholder data sharing initiatives. These examples represent possible approaches to governance that would be familiar to stakeholders. However, in the end, GDDS governance does not need to be limited by these models and can build on these approaches.

6.1.EMODnet

EMODnet ORGANISATIONAL STRUCTURE
EMODnet, is a service of the European Commission, funded by EMFAF, and owned and managed by the European Commission Directorate-General for Maritime Affairs and Fisheries (DG-MARE), who is the overall responsible entity. It is a partnership of more than 120 organisations and has no legal status on itself (it is not a legal entity). The EMODnet Secretariat, the seven EMODnet thematic groups, and the EMODnet Data Ingestion facility are operated through service contracts where the European Climate, Infrastructure and Environment Executive Agency (CINEA) is the Contracting authority. In essence this means that they are executed as projects and funded on a competitive basis with cycles ranging from two to four years, implemented by the service contractors and Member states, which build upon in-kind contributions by Member States for data collection and general data management. EMODnet is therefore a distributed network with no legal status/entity at this point in time, although this might change over time. The EMODnet Secretariat supports the coordination, communication and governance of the network and its activities/developments. The EMODnet ecosystem is composed of several layers with DG-MARE at its core, CINEA as Contracting Authority and the support from the EMODnet Secretariat for implementation. The governance layer consists of (i) the EMODnet Steering Committee comprising DG MARE, CINEA, the Secretariat, the EMODnet (thematic and data ingestion) project coordinators and a number of representatives from other Commission Services and agencies (e.g. DG DEFIS, DG ENV, DG RTD, …); and (ii) the European Commission (EC) Marine Knowledge Expert Group (MKEG) as an

independent advisory body. The (iii) partnership layer contains these "inner" layers plus all EMODnet project partners of the thematic groups and the Data Ingestion Service as well as the EMODnet Associated Partners. Finally, beyond the partners, the (iv) outer layer comprises all external stakeholders and EMODnet data providers and users.

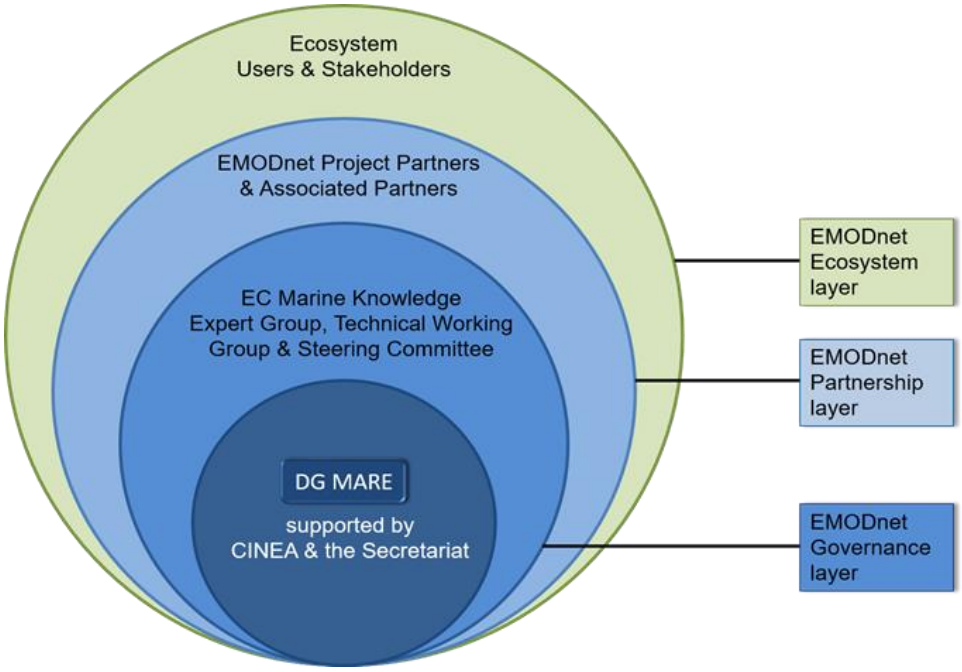

Figure 9: Governance Structure of EMODnet

EMODnet DIVISION OF ROLES

EMODnet thematic partners: EMODnet covers **7 thematic disciplines** (bathymetry, biology, geology, chemistry, physics, seabed habitats, human activities). EMODnet partners aggregate and harmonise multi-parameter datasets; make data available with searchable metadata as downloadable datasets and/or map layers or through web services (OGC). EMODnet thematic partners produce integrated data products, which are owned by the EU and are therefore published under a CC BY 4.0 licence (open data licence).

EMODnet Data Ingestion: Facilitates additional data managers to ingest their marine datasets for further processing, publishing as open data and contributing to applications for society.

EMODnet Steering Committee: It is the main governance body guiding the development of the various EMODnet activities and outputs. Currently, it comprises representatives from Contracting authority CINEA and the EC DG MARE as well as representatives from the Secretariat, the thematic assembly groups, the data ingestion service, and the Flanders Marine Institute (VLIZ), which hosts the EMODnet Central Portal. Its overarching aim is to promote coordination and consistency among the main EMODnet partners (while ensuring that their specific needs are also taken into account), the Contracting Authority, the Secretariat and DG MARE to maintain and further develop EMODnet as a performant operational fit-for-purpose user friendly marine in situ

data service aligned with the policy vision and targets set by DG MARE. The role of the EMODnet Secretariat in the Steering Committee is to chair and facilitate interactions between different members of the EMODnet Steering Committee to reach the latter's objectives as described in the Terms of Reference. The objectives include, among others: the provision of advice for the further development of the EMODnet Central Portal; the exchange of information on the progress of the various EMODnet projects and initiatives; and the assessment of the developments in the marine observations and data landscape which may affect EMODnet and interactions with other marine data initiatives.

Technical Working Group: It supports the technical implementation of the Central Portal and provides advice on technical matters. It is composed of IT developers and technical experts from the EMODnet thematic projects and the Flanders Marine Institute.

EC Marine Knowledge Expert Group (MKEG): It advises the EC on matters concerning marine knowledge, including on EMODnet as a flagship EU marine knowledge initiative, related projects (e.g. EMOD-PACE EU-China project), and European initiatives including the EC Ocean Observation " Sharing Responsibility initiative". The EC MKEG members are representatives from diverse sectors of the blue economy, the private sector, and wider marine data producer and user communities. Some representatives are also members of the EMODnet Associated Partnership Scheme, allowing for cross-fertilisation between these groups to further connect with the blue economy and wider user community.

EMODnet GREEN DEAL VALUE PROPOSITION

The work that **EMODnet** has done in terms of **standardisation** and **interoperability** of marine data supports key EU policies & strategies (e.g., Marine Spatial Planning Directive, Marine Strategy Framework Directive) and international commitments (e.g., contribution to Essential Ocean Variables). Moving forward, it will facilitate not only the use of the existing data products in support of the marine related policy objectives of the European Green Deal and wider policy objectives (e.g., SGDs), including via feeding into the development of the **EU Digital Twin of the Ocean**, and the inclusion of further relevant data categories in the future. The Green Deal Data Space opens opportunities to have the EMODnet profile raised as the marine in situ data service of the European Commission. EMODnet expects that the GDDS accepts its data publishing technologies, which are based on recognized standards.

CONDITIONS TO JOIN THE GDDS

EMODnet is in itself a data space. If the GDDS wishes to leverage and onboard EMODnet data, EMODnet Central Portal offers a harvestable data catalogue and openly accessible web services. There is no need to request permission to do any of this, as all information is made available at the EMODnet website and on GitHub. EMODnet data products are published under an open data licence, which means that no permission is used to copy, download, or use data products. Similarly, EMODnet's web services are open access, which means that there are no requirements to log in or register as a user to use and benefit from them.

6.2.EPOS RI

EPOS ORGANISATIONAL STRUCTURE

EPOS has been implemented according to a governance model in which the Executive Coordination Office (ECO) and the Integrated Core Service Central hub (ICS-C) belong to EPOS ERIC and are located inside the ERIC perimeter. ECO represents the legal seat of EPOS ERIC. An ERIC (European Research Infrastructure Consortium) is a specific legal form that facilitates the establishment and the operation of RIs with pan-European dimension. The decision body of an ERIC is its General Assembly composed of Members representing National Authorities and funding agencies.

A legal agreement (Collaboration Agreement) is signed between EPOS ERIC and the Service Provider. The Service Providers bring together Data Providers through the shared EPOS data policy.

Thematic Core Services (TCS) represent the community-specific integration and they bring in the governance framework that is needed to ensure the data and service provision (i.e., access to Data, Data products, Services and Software) within the EPOS Delivery Framework. Currently, nine TCS are formally established through the signature of a Consortium Agreement among different research organisations. TCS represent the community governance of the data generation and management, ensuring participation, the sharing of the EPOS mission, the coordination of data and service providers according to agreed data policies and access rules, as well as the community building to tackle scientific challenges and innovation. TCS oversees the services and the data provision to EPOS through designated organisations (Service Providers).

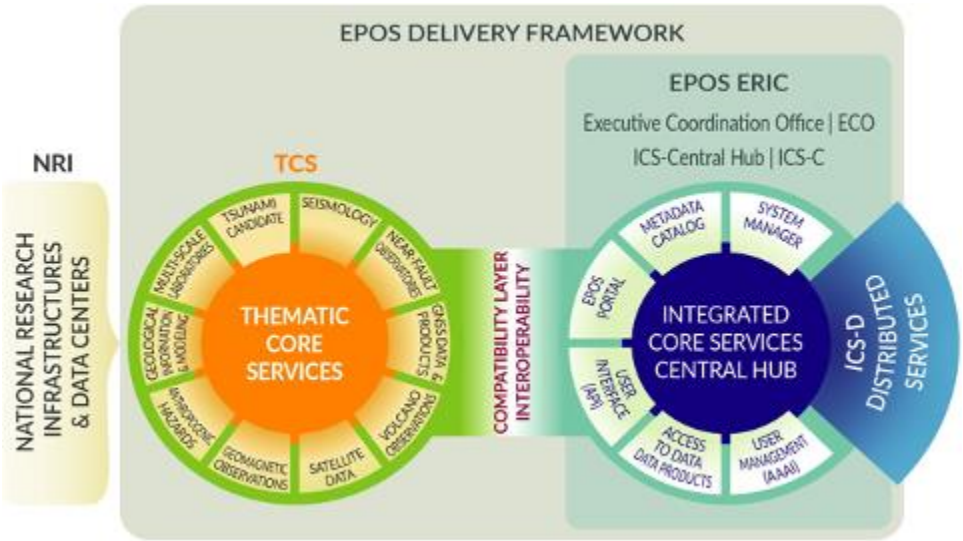

Figure 10: EPOS delivery framework

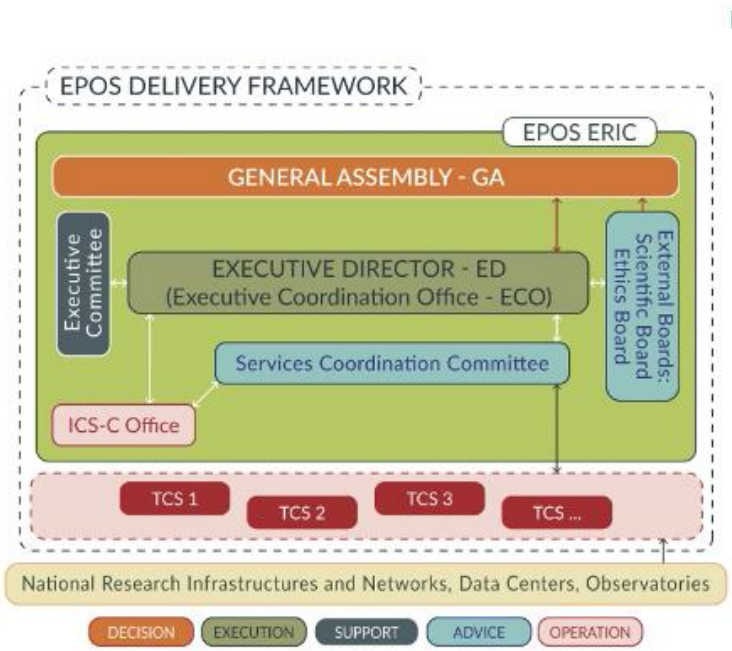

Figure 11: EPOS Governance model

EPOS DIVISION OF ROLES

Key roles include data providers, data curators, system administrators, and end-users (researchers, educators, etc.). Data providers need to ensure the availability and quality of their data, curators manage and organise the data, system administrators maintain the system, and end-users utilise the data.

The core layer of the EPOS delivery framework is managed by EPOS ERIC, which signs Collaboration Agreements for Data and Service provision with the Thematic Core Services (no Service Level Agreement). Main tasks of these agreements are:

- Provision and maintenance of the access to data through web-services based on standard protocols and allowing to search, access and download data;
- Enhancement of web-services robustness and availability to meet relevant performance specifications;
- Provision of data, metadata and access services in compliance with the FAIR data principles;
- Provision and maintenance of metadata describing TCS webservices in compliance with the EPOS metadata format;
- Provision of online documentation for web services;
- For DDSS requiring authentication and/or authorization an Authentication and Authorization system guaranteeing interoperation with OpenIDConnect and/or OAuth2 standards has to be provided.

From a financial perspective, the costs for operating the ECO and the ICS-C are supported through the provision of host contributions by hosting countries, namely Italy for the ECO, and France and

UK for the ICS-C. Host contributions are necessary to run the ECO and the ICS-C without financially impacting on the EPOS ERIC cash flow provided through the membership fees from the Members and the Observer of the ERIC. From a legal perspective, the partnership agreement signed by EPOS ERIC and the hosting organisations transforms the agreed financial and technical frameworks into an effective governance background for operating the EPOS RI.

According to the EPOS ERIC Statutes, the costs for operating the TCS, as defined in the TCS cost-book, are only partially supported by the ERIC and mainly at national level by the involved research organisations. This is an essential element of the EPOS ERIC sustainability plan, because TCS operation relies on in-kind resources provided by research organisations owning the NRIs as well as on further in kind contributions provided by national authorities (ERIC Country Members).

EPOS GREEN DEAL VALUE PROPOSITION

The EPOS community has already invested a lot of energy and time to build its own data space where Earth science researchers can exchange data through interoperable protocols and interfaces. The EPOS's members expect that the GDDS can complement their efforts in the following aspects:

- Focus on potential users and use cases goals, bringing the scientists and users to the forefront to solve specific problems.
- Providing generic services , e.g. AAI, that can expand the interoperability and the flexibility of the EPOS's framework.
- Availability of reward mechanisms applied systematically to each stakeholder (for example, publications, profit, …).
- Support with emerging technologies, e.g applying AI to data.
- Branding and exploitation of community solutions, e.g. training , upskilling

CONDITIONS TO JOIN THE GDDS

Involvement in any contractual agreements with GDDS would need technical assessment and approval from the ECO.

6.3.GOS4M

GOS4M ORGANISATIONAL STRUCTURE

GOS^{4}M is a Flagship of the Group on Earth Observation (GEO) aimed to aimed to support the Minamata Convention on Mercury Secretariat, the UN Environment Mercury Fate & Transport Partnership and all Nations in the follow up of the Conferences of Parties (COP) related to the Effectiveness Evaluation and Global Monitoring framework. The Flagship has no legal status on itself but has its organisational structure that comprises a Steering Committee (SC), a Scientific Advisory Board (SAB) and Focal Points (FPs). The Steering Committee will consist of one Representative of each Member of GOS^{4}M.

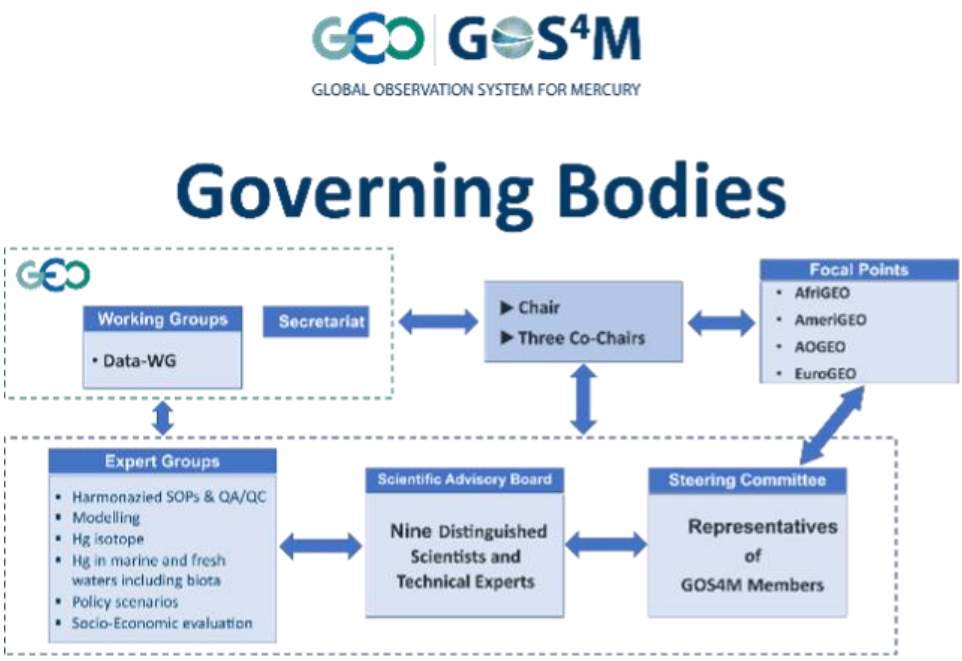

Figure 12: GOS4M Governance structure

GOS4M DIVISION OF ROLES

The SC consists of one Representative of each Member of GOS^{4}M. It will:

- ensure the efficient management and implementation of the GOS^{4}M Business Plan (BP) [referring to the up-to-date GEO Work Plan];
- revise the BP by considering the suggestions that may be provided by its members and by the SAB;
- liaise with participating organisations and institutions supporting the gathering and collection of mercury data and information;
- ensure efficient communication and outreach activities;

- coordinate the sharing of up-to-date information provided by GOS^{4}M Members with all interested Parties;
- ensure that GOS4M portal provides state-of-the-art information, data and tools in support of Parties of the Minamata Convention on Mercury;
- promote the development of joint cooperation activities and projects among its members and between its members and other organisations;
- peer-review the reports and guidance documents produced by GOS^{4}M;
- report to GEO Secretariat on the progress of GOS^{4}M and its major achievements.

The SAB is a subsidiary body composed of distinguished scientists and technical experts covering different domains of fundamental and policy-oriented research. It is established to advise the SC on matters relating to current and future mercury science and technology information. The SAB is composed of nine distinguished, well recognised experts covering one or more mercury research and policy domains who have provided a significant contribution to advancement of science and environmental policy related to global mercury pollution issues including human health. The SAB can establish an Expert Group (EG) to cover expertise not available among its members. The EG would comprise well known experts on emerging topics that might be relevant for the GOS^{4}M activities.

In order to ensure close cooperation between the GOS^{4}M activities and the Regional GEOs, FP of the GOS^{4}M community will be identified, appointed and serve as liaisons to the Regional GEOs. Their role is to facilitate communications, information and knowledge sharing, and identifying regional priorities or needs for the GOS^{4}M SC to consider. They would also advocate for GOS^{4}M within the region. The GOS^{4}M Steering Committee will establish FPs for each Regional GEO. The FPs members will also link GOS^{4}M and Regional GEO activities by facilitating the exchange of knowledge between the two groups to help inform and develop regional policy needs that can be implemented into the GOS^{4}M Knowledge Hub as part of the overall GEO Knowledge Hub.

GOS4M GREEN DEAL VALUE PROPOSITION

To support the demonstration that the Earth Observation domain can play a much more significant role in the effectiveness evaluation of policy frameworks.

Provide trusted solutions based on digital infrastructures and artificial intelligence to facilitate evidence-based decisions. GOS^{4}M has a well-established data governance for environmental information (i.e. meteorological and chemical parameters) collected from in-situ monitoring stations. No structured health datasets are available to support assessment of pollution impact. The action should foster links with relevant initiatives on health data collectors.

CONDITIONS TO JOIN THE GDDS

GOS^{4}M can join the GDDS through both stages:

- Onboarding of participants: a representative of the GOS^{4}M can be part of the GDDS stakeholder to bring needs and interests;
- Onboarding of data/services: GOS^{4}M has a harvestable data catalogue and openly accessible web services (https://sdi.iia.cnr.it/gos4mcat). There is no need to contact the Secretariat to do any of this, all information is made available on our website. GOS^{4}M data

products are published under an open data licence, no permission is used to copy, download, use data products. Similarly, GOS^{4}M's web services are open access, there's no requirement to log in or register as a user to use them.

7. Conclusion and Next Steps

This document presents a comprehensive governance framework for data spaces – both generic and specific to the Green Deal, building on best practices from a range of relevant activities, including both the data space community as well as experience in the creation and operation of digital platforms, and effective IT and cybersecurity practices and governance.

The GREAT project's five Phase 1 use cases have provided initial insights into the specific governance requirements that apply to their use cases. With this presentation of a complete generic governance framework, more detailed consultations will be undertaken in Phase 2, involving the initial five use cases, as well as five additional use cases, to determine in which areas generic requirements are appropriate, and where sector-specific requirements must be defined.

Both the generic and sectoral governance frameworks highlight how important it is to define the community served by the data space and to establish the objectives and purpose of a data space – either to achieve the goals of a single use case, or to enable a suite of use cases. This document has postulated an initial set of goals and objectives for the Green Deal Data Space, which will be refined and elaborated in consultation with the Green Deal Community of Practice and specifically with the ten use cases being used by the GREAT project as references for our preparatory efforts.

The presented generic and sectoral frameworks have been constructed within a specific contextual view of the data space landscape, which helps to align existing data sharing initiatives and community aspirations with the possibilities afforded by a common European Data Space. In addition to consulting with its Community of Practice and Reference Use Cases, GREAT will work with sister data space preparatory actions to share insights and potentially develop a broader consistent view the "data space landscape" and likely evolution of efforts over time. These discussions will address: governance, contractual agreements, common functions, standards, and possibly "metadata about Data Spaces".

More broadly, GREAT will work with projects such as the Data Space Support Centre to collect guidance in the specific areas itemized above, in order to refine our framework to be more robust.

In addition to further consultation on governance requirements, the GREAT project will translate some of the insights developed in this work package and deliverable into structures and approaches for an implementation roadmap for the Green Deal Data Space. The work of this deliverable will inform the roadmap by helping to stage development efforts through progressive "elevation" of the GDDS' objectives – moving from lower-level objectives to higher level objectives over time and helping to outline appropriate development timelines. In addition, the concept of data space initiatives, as an evolutionary step supporting progressively increasing harmonisation and consolidation.

8. APPENDIX I: Analysis of Horizontal EU Legal Framework for Data Spaces

The scope of our discussion is the European Union, but in practice the "data transactions" that are to be enabled by a data space are transactions between two parties that are each governed by the laws of one or more Member States.

Without a legal framework from a particular data space, such transactions would occur within two possible legal contexts: either that of a specific legal agreement between the parties, which usually includes a "choice of law" or similar clause specifying the jurisdiction governing the agreement, or, without such an agreement, the legal jurisdictions that govern each of the parties. In all cases the relevant jurisdiction would be one or more Member States, rather than the EU itself, so it is important to consider Member State legislation rather than just legal and regulatory arrangements (or "templates") at the EU level.

A data space is similarly governed, not by EU law *per se*, but by the versions of EU legislation that are enacted ("transposed") in the Member State whose laws are chosen to govern that data space. Since those laws also operate in conjunction with other applicable legislation in that Member State, including general commercial legislation, consideration must also be given to laws and regulation addressing legal entities formed in that country, commercial transactions, as well as consumer protection where private citizens are involved (e.g. notably as the subjects of personal data).

8.1. EU Legal and Regulatory Context

Although introduced at different times, collective legislation and regulation from the European Union (the European Union "*acquis*") establishes a foundation for data sharing across several related areas:

- Cybersecurity
- Data Privacy and Protection
- Data Access and Use
- Data Transactions: Data Intermediaries and contracts.

These areas approximately align with the progressive levels of control on data:

- Secure (and invisible)
- Visible (to one user, authenticated users, or the public)
- Findable (searchable using one or more tools)
- Accessible (viewable, downloadable)
- Usable/Interoperable and Re-usable.

8.1.1. Cybersecurity

Cybersecurity is a common foundational requirement for any digital system. As explored by the ISO 27000 standards[115], cybersecurity combines process, people, and technology to achieve the joint cybersecurity objectives of confidentiality, integrity and availability of the services and data provided or managed by an information system.

Various EU laws and regulation require that organisations ensure the cybersecurity of the information systems they operate, either because those systems hold personal or customer data, or because those systems play an important role in delivering critical services (such as energy infrastructure).

For customer or personal data, the General Data Protection Regulation (GDPR), Payment Services Directive 2 (PSD2), and proposed ePrivacy Regulation all require information system operators to ensure the safety of the customer or personal data held by those systems.

For critical infrastructure, the NIS (Network and Information Systems) Directive is an EU-wide legislation that aims to improve cybersecurity in critical infrastructure sectors, such as energy, transportation, and healthcare – and some pan-European data spaces might be designated as critical infrastructure.[116] including compliance to cybersecurity requirements. The directive requires member states to establish national strategies for the security of network and information systems and to designate competent authorities to oversee their implementation. The NIS Directive also requires operators of essential services and digital service providers to take measures to manage cybersecurity risks and report significant incidents.

Cybersecurity requirements are explored further under "Digital Platform Governance" (section 2.2) below.

8.1.2. Data Privacy and Protection

Data privacy and data protection build on cybersecurity concepts. Cybersecurity ensures that unauthorised parties cannot access protected data, while data privacy addresses how and to what extent parties should have access.

The principles of personal data privacy and protection have been advanced by the European Union since 2014. The idea of more general rights in data, including non-personal data, took form in the Commission's 2017 Communication 'Building a European Data Economy' which introduced the notion of a '*data producer's right'* to protect industrial or machine-generated data.

The General Data Protection Regulation (GDPR) is the principal regulation implemented by the European Union (EU) that ensures privacy and protection of personal data. The GDPR came into effect in 2018, and it outlines strict rules for how personal data should be collected, processed,

[115] ISO 27000 outlines the security techniques necessary to properly safeguard customer data. Organisations implement the requirements outlined in ISO 27000 standards and verify the effectiveness of their ISMS through an ISO 27001 audit.

[116] These initiatives are also subject to Article 12.6/12.5 of the (EU) 2021/694 regulation in the implementation of the Digital Europe program, ensuring duly justified involvement of third party countries in the activity,

and stored. It also establishes key principles for access and use or personal data which are outlined in the next section.

In addition to the GDPR, there are several other EU laws and regulations that affect personal data privacy and protection, including:

- ePrivacy Regulation: The ePrivacy Regulation is a proposed regulation that aims to strengthen the protection of privacy and confidentiality in electronic communications. It would complement the GDPR by providing specific rules on the use of electronic communications data, such as metadata, cookies, and tracking technologies.
- PSD2: The Payment Services Directive 2 (PSD2) is an EU directive that regulates payment services and providers in the EU. The directive aims to enhance competition and innovation in the payment industry while ensuring a high level of security for users' data and transactions. PSD2 introduces new requirements for strong customer authentication (SCA) and open banking.
- Schrems II ruling: The Schrems II ruling is a recent decision by the European Court of Justice (ECJ) that invalidated the Privacy Shield agreement between the EU and the US. The ruling found that the Privacy Shield did not provide adequate protection for EU citizens' personal data when it was transferred to the US. The ECJ also reaffirmed the importance of the GDPR's data protection principles and requirements for international data transfers, which require data controllers to ensure adequate safeguards are in place to protect data when it is transferred outside of the EU.

8.1.3. Data Access and Use

When a potential data use and user are identified, the corresponding mechanisms for data access must be defined before that use, by that user, can be allowed. As EU legislation on this topic expands, so do the specific cases and required mechanisms, which are summarised below. Given the variety of these circumstances, related mechanisms, and requirements, extending W3C's Data Protection Vocabulary may be an effective approach to properly labelling any data that have particular access and use mechanisms and capturing related metadata required to implement those mechanisms (e.g. identifying data subjects and data controllers).

8.1.3.1. Personal Data

The GDPR defines various mechanisms for individuals to control access to and use of their personal data. It grants them the right to know what personal data is being processed, to have that data deleted, and to object to the processing of that data. The GDPR also imposes significant fines on organisations that violate these rules. Other EU laws and regulations mentioned above also address the access to and use of personal data. The ePrivacy Regulation gives users more control over how their data is used, particularly in the context of online advertising. The Payment Services Directive 2 (PSD2) gives users more control over their financial data and enable them to access a wider range of payment services.

8.1.3.2. Data Held by the Public Sector

8.1.3.2.1. INSPIRE Directive

The INSPIRE directive 2007/2/EC entered into force in 2007 laying down general rules to establish an Infrastructure for Spatial Information in Europe, for the purposes of EU's environmental policies and policies or activities which may have an impact on the environment.

INSPIRE should build on infrastructures for spatial information established and operated by the Member States (Article 1).

Since its inception, Member States (MS) have made available tens of thousands of datasets via interoperable metadata and services in line with the directive[117] becoming one of the biggest efforts in Europe to harmonise spatial data infrastructures.

The directive binds public authorities at all levels in the MS legally via the Implementing Rules, a set of specifications at the abstract level with requirements on the provision of data/metadata, network services, data sharing, interoperability and monitoring and reporting. The Implementing Rules apply to the following specific requirements described above:

- **Metadata** Implementing Rules[118]: reflected in **GGF-4.1.02/DG/F** Define Metadata/Self-Descriptions required for all data.
- **Data Specifications** Implementing Rules[119]: "Specify common data models, code lists, map layers and additional metadata on the interoperability to be used when exchanging spatial datasets", reflected in **GGF-3.1.01/DS/F** Domain Data Models and Interoperability Standards.
- **Network Services**[120], **Data and Service Sharing**[121], **Spatial Data Services**[122]: These three sets of implementing rules are reflected in the array of services to be offered through a Data Space, categorised by type:
    - **GGF-3.1.02/DS/F: I**ntermediation/ Marketplace/ Catalogues
    - **GGF-3.1.03/DS/F:** Ancillary Services: Data Preparation, Encryption, Anonymization, Transformation.
    - **GGF-3.1.04/DS/F:** Enrichment, Aggregation, Fusion, Analysis, AI/ML.

To support the rules a set of non-binding Technical Guidance documents provide information of how Inspire can be implemented with the adoption of standards such as those from the Open Geospatial Consortium (OGC).

The INSPIRE directive is under review, together with the Directive 2003/4/EC on public access to environmental information, under the initiative GreenData4All[123] to bring them up to date with the Green Deal strategic actions. Both initiatives represent the "backbone of the environmental information management covering the whole of EU environmental policy"[124] (Kotsev et al).

As of late summer 2023, the GreenData4All initiative will shortly begin a public consultation, and the GREAT project will provide input to this consultation.

8.1.3.2.2. Open Data Directive

The Open Data and Public Sector Information Directive EU 2019/1024 entered into force 16 July 2019 replacing the preceding Public Sector Information Directive 2003/98/EC. The directive aims to ensure that public sector bodies make information available and reusable for the benefit of

[117] https://inspire-geoportal.ec.europa.eu/overview.html?view=thematicEuOverview&theme=none
[118] https://inspire.ec.europa.eu/Legislation/Metadata/6541
[119] https://inspire.ec.europa.eu/data-specifications/2892
[120] https://inspire.ec.europa.eu/network-services/41
[121] https://inspire.ec.europa.eu/data-and-service-sharing/62
[122] https://inspire.ec.europa.eu/spatial-data-services/580
[123] 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

[124] https://publications.jrc.ec.europa.eu/repository/handle/JRC126319

citizens, businesses, and the overall society. The directive promotes the use of open data to stimulate innovation in products and services while ensuring transparency and fair competition in the internal market. It applies to public sector bodies such as government departments, agencies, ministries, local authorities, public institutions, and organisations that are funded by public money which may include universities. It also applies to public undertakings e.g., such as public utilities and research data produced with public funding that Member States make available under national open access policies and compatible with FAIR principles.

Public sector bodies are encouraged to create data based on the principle "open by design and by default" (Article 5) promoting licences that allow for broad reuse, including commercial exploitation, free of charge or at minimal cost of reproduction and without necessary restrictions while respecting privacy and intellectual property rights.

The directive applies to "documents" defined as content in any medium including paper or electronic form such as sound, visual or recordings. Documents (and their metadata) should be made available in formats that are open, machine-readable, accessible, findable and re-usable.

The directive introduces the concept of High Value Datasets (HVDs) defined as "documents held by a public sector body, the re-use of which is associated with important benefits for the society, the environment and the economy". Public sector organisations must make HVDs available free of charge, in machine-readable format, via Application Programming Interfaces (APIs) and, where relevant, as a bulk download. Annex I of the Directive lists the HVDs in the following categories: geospatial, earth observation and environment, meteorological, statistics, companies and mobility.

The Open Data Directive required adoption by the Commission of an implementation act specifying a list of HVDs. On 22 of December 2022 the Commission adopted the Implementing Act on a list of High-Value Datasets 2023/138, which shall apply from June 2024. Under this regulation, public sector bodies holding high-value datasets listed in the Annex shall ensure that these are made available in machine-readable formats via APIs corresponding to the reasonable needs of re-users. It also includes reporting obligations every two years specifying the measures taken to implement the regulation.

8.1.3.2.3. Data Governance Act

The Data Governance Act (DGA) addresses *inter alia* mechanisms for access to public sector data that is subject to legal restrictions and was therefore left out of the scope of the 2019 Open Data Directive. Specifically, the DGA covers public sector data that is legally protected on the grounds of (a) commercial confidentiality including trade secrets; (b) statistical confidentiality; (c) intellectual property rights of third parties; (d) protection of personal data.

The DGA does not specify any rights of access or use for this data, but if a public sector body decides that access and use are to be allowed, the DGA requires that it:

- ensures the preservation of the data's protected nature via appropriate technical and organisational safeguards (e.g., anonymization of the data or the provision of a secure processing environment for data access and re-use);
- imposes confidentiality requirements on data re-users;

- ensures that non-personal confidential data or data protected by intellectual property rights are transferred to third countries only under appropriate safeguards and possibly with the aid of model contractual clauses adopted via implementing act by the European Commission.

and prohibits public sector bodies from:

- exercising the *sui generis* database right to prevent or restrict data re-use;
- entering into exclusive arrangements for the re-use of protected data held by these public sector bodies.

In contrast to the Open Data Directive, the DGA allows for data re-use to be made contingent upon a fee, provided the fee mechanism is transparent, non-discriminatory, proportional, objectively justified, and not anti-competitive. Incentives are to be created for the re-use of data for non-commercial purposes, as well as its re-use by SMEs and start-ups.

All these requirements of the DGA establish a pro-active responsibility on the part of public sector bodies to enable the re-use of data they hold. Such bodies are now expected to provide and arrange for a secured environment for data re-use, to arrange the anonymization or pseudonymization of personal data or delete commercially confidential information, as well as to supervise the re-use of data (by the data re-users) in order to prevent re-use detrimental to the rights of third parties and to ensure that confidential information is not disclosed as a result of the re-use. Rather than merely curating data (both technically and legally) with a view to its re-use, public bodies are now also expected to *tailor* such activities to the *specific* data re-use and data re-user's purpose for data processing. They are expected to accommodate specific the rights and legitimate interests of protected third parties on the one hand (data subjects and holders of intellectual property rights or trade secret rights) and data re-users on the other hand.[125]

Note that public sector data includes research data produced with public funding, so the requirements of the DGA apply fully to research data.

8.1.3.3. Data Altruism

In addition to clarifying mechanisms for access and use of public sector data, the Data Governance Act establishes a new category of data access, namely "data altruism", in which one or more legal or natural persons may choose to make their personal or non-personal data (respectively) available "for objectives of general interest as provided for in national law, where applicable, such as healthcare, combating climate change, improving mobility, facilitating the development, production and dissemination of official statistics, improving the provision of public services, public policy making or scientific research purposes in the general interest".[126] Data altruism is possible in some Member States through existing national law, and natural and legal persons may engage in data altruism outside the scope of the DGA.

The DGA extends this activity through two mechanisms:

[125] EUH4D D3.7 Evaluation and recommendations on the legal conditions for trading data in a complex ecosystem https://cordis.europa.eu/project/id/951771/results
[126] DGA R16

- Encouraging Member States to support data altruism through "organisational or technical arrangements, or both, which [...] could include the availability of easily useable tools for data subjects or data holders for giving consent or permission for the altruistic use of their data, the organisation of awareness campaigns, [...] a structured exchange between competent authorities on how public policies, such as improving traffic, public health and combating climate change, benefit from data altruism, [and] establish[ment of] national policies for data altruism."[127]
- Definition of a new entity, a "Data Altruism Organization" (DAltO), that must be a not-for-profit organisation established in the EU (or with legal representation within the EU), complying with transparency requirements, and having specific safeguards in place to protect the rights and interests of data subjects and data holders.

The DGA contemplates that Data Altruism Organizations would provide data storage and processing services for the data entrusted to them, thereby creating data repositories, and contributing to the creation of "data pools" supporting objectives of general interest. DAltOs must register with designated competent authorities in the Member State in which they are established and would need appropriate internal operating procedures to correctly manage the data in their trust, meeting transparency requirements and complying not only with the DGA's requirements, but also the GDPR (for personal data), as well as a planned "Rulebook" establishing additional or modified requirements for DAltOs.

8.1.3.4. Data from Connected Products

The EU's proposed Data Act (DA) addresses data collected or generated by "connected products" (which are broadly defined, ranging from cars to smart phones to data collected by Internet-of-Things devices):

- the DA creates a right of access to this data for the product's lawful users. Access to any personal data where the identified data subject is not the lawful user is subject to appropriate requirements under GDPR.
- The DA prohibits the use of any personal data collected by the connected product by its manufacturer unless this is covered by a contractual agreement with the lawful user, and in any event the manufacturer may not allow the data collected (personal or otherwise) to be used to "derive insights [...] that could undermine the commercial position of the user". (This requirement assumes the user is not a natural person.)

8.1.3.5. Data Held by a Business Legally Required to Provide its Data to another Business

Any business required under any EU law to make the data it holds available to another business, including the manufacturer or operator of the connected products, is required by the proposed DA to do so based on a contractual agreement for access and use, on a fair, reasonable, non-discriminatory, transparent and non-exclusive basis.

[127] DGA R45

8.1.3.6. Data Requested by the Public Sector in Exceptional Circumstances

Under the proposed DA, data holders may be asked to provide public sector bodies with access to data they hold in a variety of circumstances, such as responding to public emergencies, preventing or recovering from a public emergency, or fulfilling a legally required function for which the holders' data is required because no alternative data can be made available with reasonable effort. Subject to due process governing such requests, the data provided in this way to public sector bodies is not covered by the general requirements applying to public sector data and must be handled according to specific instructions.

8.1.4. Data Transactions

8.1.4.1. Data Intermediaries

A key concept introduced by the Data Governance Act (DGA) is that of "data intermediaries" and "data intermediation services". These services are expected to figure prominently in data spaces and are also expected to place significant technical and administrative burdens on data spaces, so the related requirements are discussed here in detail.

A 'data intermediation service' is a service that links data subjects and data holders with data users, through commercial relationships that might involve technical, legal or other means. Examples of data intermediation services include "data marketplaces [through] which undertakings could make data available to others, orchestrators of data sharing ecosystems that are open to all interested parties, for instance in the context of common European data spaces [emphasis added], as well as data pools established jointly by several legal or natural persons with the intention to license the use of such data pools to all interested parties in a manner that all participants that contribute to the data pools would receive a reward for their contribution".[128]

"Data cooperatives" are defined as one type of data intermediary. They are defined 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, including with regard to making informed choices before they consent to data processing, to exchange views on data processing purposes and conditions that would best represent the interests of its members in relation to their data, and to negotiate terms and conditions for data processing on behalf of its members before giving permission to the processing of non-personal data or before they consent to the processing of personal data".[129] Data altruism organisations are excluded from being data intermediaries, so a data cooperative is different from a data altruism organisation.

The definition of data intermediation services excludes[130]:

- services that aggregate, enrich or transform data for the purpose of adding value and licence the resulting data to data users, without establishing a commercial relationship between original data holders and ultimate users of the value-added data;
- services that intermediate copyright-protected content;

[128] DGA R28
[129] DGA R15
[130] DGA Article 2(11)

- services exclusively used by one data holder or by multiple legal persons in a closed group (including supplier or customer relationships or collaborations established by contract, in particular those primarily involving the Internet of Things);
- data sharing services offered by public sector bodies that do not aim to establish commercial relationships.

Other undertakings excluded from the definition of data intermediaries include[131]:

- Recognised data altruism organisations, unless they aim to establish commercial relationships between data subjects and/or data holders and data users.
- Other not-for-profit entities whose activities consist of seeking to collect data for objectives of general interest, made available by natural or legal persons on the basis of data altruism, unless they aim to establish commercial relationships between data subjects and/or data holders and data users.

The second exception seems to refer to, for example, repositories that aim to enable the re-use of scientific research data in accordance with open access principles[132].

Data intermediaries performing data intermediation services must submit a notification of these activities to competent national authorities. They must be a legal entity established in the EU (or with legal representation within the EU), must allow the competent authority in their jurisdiction to monitor their activities, and must meet the following conditions[133]:

- data associated with its services must be put at the disposal of data users, and the data intermediation services must be provided through a separate legal person;
- the commercial terms, including pricing, for the services cannot be "bundled" with other services;
- data associated with its services can only be used for the development of that data intermediation service, for example for the detection of fraud or cybersecurity, and shall be made available to the data holders upon request;
- data formats will be converted only to enhance interoperability, if requested by the data user, mandated by law, or to ensure harmonisation with international or European data standards;
- additional specific tools and services may be offered to data holders or data subjects for the specific purpose of facilitating the exchange of data, such as temporary storage, curation, conversion, anonymisation and pseudonymisation, such tools being used only at the explicit request or approval of the data holder or data subject;
- procedures for access to its service are fair, transparent and non-discriminatory, regarding both prices and terms of service;

[131] DGA Article 15
[132] DGA R29
[133] DGA Article 12

- procedures must be in place to prevent fraudulent or abusive practices in relation to parties seeking access through its data intermediation services;
- in the event of its insolvency, the provider shall ensure a reasonable continuity of its services and, where data is stored, shall allow data holders and data users to retrieve their data and allow data subjects to exercise their rights;
- ensure reasonable interoperability with other data intermediation services, inter alia, by means of commonly used open standards in its sector of operation;
- adequate technical, legal and organisational measures are in place to prevent unauthorised transfer of or access to non-personal data;
- promptly inform data holders in the event of an unauthorised transfer, access or use of the non-personal data that it has shared;
- ensure an appropriate level of security for the storage, processing and transmission of non-personal data, as well as the highest level of security for the storage and transmission of competitively sensitive information;
- services offered to data subjects shall be in the data subjects' best interest, in particular by informing and advising data subjects in a concise, transparent, intelligible and easily accessible manner about intended data uses by data users and standard terms and conditions attached to such uses before data subjects give consent;
- where tools are provided to obtain consent from data subjects or permissions to process data, specify the third-country jurisdiction in which the data use is intended to take place and provide data subjects with tools to both give and withdraw consent and data holders with tools to both give and withdraw permissions to process data;
- maintain logs of the data intermediation activity.

For a data space, most of the DGA's requirements for data intermediaries, assuming they apply to the "organising entity" of the data space, seem to represent best practices. However, several aspects bear further examination:

- Separating the provision of data intermediary services from the provision of other value-added services may be difficult and would certainly require careful planning in advance of the design of those services.
- Exclusion of intermediary services provided to a closed group may be an easily accessible "exception" since most data spaces are considering the use of standard framework agreements that would govern access to the data space, which would create such a closed group.
- Many data spaces are considering mechanisms for data exploitation implicitly creating commercial relationships, immediately invoking the requirements placed on a data intermediary. Data spaces focussing on open access data (from public sector bodies and publicly funded research), might be able to avoid the requirements for data intermediaries. However inclusion of any data being offered commercially (on whatever terms) may trigger

data intermediary treatment. Similarly, any mechanisms for charging fees for the use of a data space, to support its sustainability in general, and to fund core service providers specifically, might also trigger such treatment.

It seems likely that most services that enable data transactions would be treated as data intermediation services, and their providers would need to be legal entities complying with the Data Governance Act.

8.1.4.2. Contractual Agreements for Data Provided to Micro or SMEs

Per the proposed DA, contractual agreements to provide access to and use of data to any micro, small or medium-sized enterprise cannot be unilaterally imposed (e.g. through a standard, non-negotiated contract) and cannot contain any provisions deemed unfair according to a number of listed criteria.

8.1.4.3. Large Online Platforms

The Digital Services Act (DSA) and the Digital Markets Act (DMA) were proposed by the European Commission in December 2020, with the objective of regulating “large online platforms”, such as social media and e-commerce websites, to ensure they do not engage in anti-competitive practices, and to protect users' fundamental rights online. The DSA and DMA include provisions related to data protection and privacy, such as transparency requirements for online advertising, and rules for how platforms should handle illegal content, such as hate speech or terrorist propaganda. Given the proposed definition of a “large online platform” (e.g. having more than 10,000 business customers in the EU) it may be possible that a data space would fall under the jurisdiction of the DSA and DMA. Similarly, given the objective of the DSA and DMA to limit anti-competitive practices, any data space seeking to avoid duplication of data and services in a particular sector might be seen as limiting competition and might therefore be governed by the DSA or DMA.

9. APPENDIX II: Best Practices in Governance of Multi-Stakeholder Alliances and Social Enterprises

Relevant best practices in governance can be identified from two domains: multi-stakeholder alliances, and social enterprises.

9.1. Multi-Organization Alliances

Multi-organizational alliances are not limited to the technical domain, thousands of such organisations exist around the world[134], and a number of best practices for their governance have been identified.

Van den Broek and Van Veenstra (2015)[135] compile four archetypical modes of governance among the different organisations within a given alliance focussing on "data collaboration":

- Market: governed by formal contracts, with little focus on trust. This might be thought of as the "base case" for alliances -- instrumented through specific bilateral agreements, without any overarching governance structure.
- Bazaar: focussed on collaborative activity, based on and building on reputation. Today this might be best illustrated by the "gig" economy, or by the project-oriented alliances created during film production.
- Hierarchy: marked by administrative focus and bureaucratic approach. Supply chains, centred around the dominant manufacturer, are a good example of a hierarchical alliance. While contracts are in place, the full range of relationships between each party is not governed by contracts, but instead by the administrative requirements of the dominant partner.
- Network: relying on common goals, social contract and reciprocity. Cooperatives and associations are common networked forms of organisations.

Within the network mode of governance, additional organisational options for alliances are:

- Choice of organisational form for the central entity: dictated by jurisdictional scope and its legal ability to accommodate the governance structure desired.
- Choice of membership structure: both "single stakeholder" and "multi-stakeholder" models of governance model are possible (referring to different categories of constituents (e.g. doctors, vs. nurses, vs. patients). As discussed in the next section, where the alliance has a "social aim" research indicates that multi-stakeholder governance encourages the building of trust, knowledge and learning among the broader group of stakeholders, which in turn contributes to the success of the alliance itself.

Various governance organs and mechanisms can be used to structure stakeholder engagement

[134] The most notable examples of alliances are cooperatives: in 2017 there were 1,420 co-operatives across 52 countries with a turnover of more than US$100 million, and the largest 300 cooperatives had an average combined turnover of over US$3 billion each._Other examples of alliances are associations -- typically alliances of individuals, often to further the professional stature and development of their members. Many professional associations are very large in their own right: The Institute of Electrical and Electronic Engineers (IEEE) is a US non-profit corporation with over 400,000 members worldwide and annual turnover of over US$500 million. Industry associations are also significant. Birchall, Johnston. (2017). The Governance of Large Co-operative Businesses. https://www.uk.coop/sites/default/files/uploads/attachments/governance-report2017finalweb.pdf

[135] van den Broek, Tijs and van Veenstra, Anne Fleur, "Modes of Governance in Inter-Organizational Data Collaborations" (2015). ECIS. 2015 Completed Research Papers. Paper 188. ISBN 978-3-00-050284-2. http://aisel.aisnet.org/ecis2015cr/188

and consultation, as well as ultimate decision-making, even in single stakeholder governance structures:

- Organizations with large numbers of "ultimate" members sometimes create an intermediate member council, to which members are elected, and which then elect a governing board or council.
- Organizations requiring geographic or constituent balance can create multiple corresponding intermediate councils (e.g. regional councils, user councils) which can then elect their own representatives to the governing board. IEEE's governance structure uses a rich set of intermediate councils to ensure representative balance among a very large membership.
- Even when intermediate councils do not formally elect representatives to the board, they can act as advisory bodies, whose advice the board should consider as part of the governance framework.

Regardless of structure, the use of deliberative decision-making, supporting open communication and consensus building, is widely identified as a best practice.

Some organizations (including for profit organizations) apply various best practices to make better decisions:

- They seek individuals with specific expertise and competencies to serve as directors or governors.
- They specifically instruct their decision-makers (directors or governors as well as senior leaders) to act in the best interest of the organization rather than in the interest of any parties they might represent. For example, large shareholders may have the ability to select board members, who may find that the interests of the organization and of the shareholder that placed them on the board are different.
- They require explicit declarations of any conflicts of interest by each board member.
- They require some number or proportion of independent directors on the board or governing council, to ensure expertise is included, to encourage the use of best practices, and to avoid "control" of governance processes by dominant members.

Alliances employ the following common operational structures or characteristics:

- Specialized resources focussed on making the alliance work, e.g., personnel assigned to the work of the alliance, formal secondment, creation of a separate entity.
- Formalized procedures (documented, standardized) to improve its effectiveness. The Boundary Resources identified above as a key dimension of technological platforms would be an example of such a formalized procedure.
- Communications mechanisms that support a broad set of interfaces between and among the organizing entity, members and stakeholders, typically through working groups and committees.
- Monitoring and evaluation mechanisms to track performance of the alliance in meeting its objectives.

Macdonald, et al. (2019)[136] found that alliances with robust mechanisms both for communications and for monitoring and evaluation, experience improved resiliency and organizational capacity, implicitly contributing to greater effectiveness and impact. At the same time the number of partners in the alliance, as well as their diversity, detracted somewhat from that effectiveness, suggesting that alliances need to find the right balance between stakeholder inclusiveness and relevance to optimize both engagement and effectiveness.

9.2.Social Enterprises

Social enterprises (SEs) harness entrepreneurial dynamics to create public goods and serve the public or general interest. Since some data spaces are expected to create public benefits, governance aspects of SEs will be important for such data spaces. The "data altruism organizations" and "data cooperatives" contemplated by the Digital Governance Act fit the SE model.

Sacchetti et al. (2019)[137] summarize several key organizational features of SEs:

- Agreement on the "social aim" that the SE seeks to achieve.
- Inclusive and participatory approach embodied in governance and decision-making processes. This principle applies to inclusion of stakeholders in the governance process, rather than just owners, shareholders or founders.
- The "redistribution" (accumulation and reinvestment) of surplus resources. This can refer simply to building a reserve fund to ensure the sustainability of the organization, or to active collection and redistribution of surplus resources[138].
- Fulfilling non-monetary motivations of participants through the first three features: participants include both individual employees (and volunteers) as well as the organizations participating in the enterprise.

The commitment to serve a higher purpose and to create benefits for more than just the primary stakeholders of an organization (called the "cooperative pact" by Sacchetti et al.) can motivate stakeholders to subsume their direct interests to that larger purpose.

Several studies[139][140][141] examine the relationship of stakeholding and governance to the effectiveness of a SE in fulfilling its mission, and all identify advantages for multi-stakeholder

[136] Macdonald, Adriane & Clarke, Amelia & Huang, Lei. (2019). Multi-stakeholder Partnerships for Sustainability: Designing Decision-Making Processes for Partnership Capacity. Journal of Business Ethics. 160. https://doi.org/10.1007/s10551-018-3885-3

[137] Sacchetti, S., Borzaga, C., Tortia, E. (2019) "The institutions of livelihood and social enterprise systems", Euricse Working Paper Series 109 | 19. https://papers.ssrn.com/sol3/papers.cfm?abstractid=3519810

[138] For example, for a cloud federation, the sharing of infrastructure resources might be regarded as a type of redistribution in support of the federation's mission. For a data federation, creating structures to capture and share value created through the contributions of multiple partners would be an example of such "redistribution".

[139] Borzaga, Carlo & Mittone, Luigi. (1997). "The Multi-Stakeholders Versus the Nonprofit Organisation". Discussion Paper from Universita degli Studi, Trento, Italy. https://www.researchgate.net/publication/24136644TheMulti-StakeholdersVersustheNonprofitOrganisation

[140] Fazzi, Luca. (2012). "Social Enterprises, Models of Governance and the Production of Welfare Services". Public Management Review 14. 359-376. https://doi.org/10.1080/14719037.2011.637409.

[141] Sacchetti, S. & Borzaga, C. (2017), The Foundations of the "Public" Organisation: Strategic Control and the Problem of the Costs of Exclusion, *Euricse Working Papers,* 98|17.

structures. Despite these advantages, Sepulveda et al. (2020)[142] recognize that effective engagement of multiple types of stakeholder in both stewardship and decision-making requires purposeful action, including adoption of suitable legal forms, inclusive organisational cultures, visionary leadership and concrete actions that align with the organisation's social mission: "it is neither structure nor culture, but rather a synergistic interplay of the two that matters".

[142] Sepulveda, Leandro & Lyon, Fergus & Vickers, Ian. (2020). Implementing Democratic Governance and Ownership: The Interplay of Structure and Culture in Public Service Social Enterprises. VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations. https://doi.org/10.1007/s11266-020-00201-0.

10. APPENDIX III: Best Practices in Governance of Information Technology Activities as well as Cybersecurity

10.1. IT Governance

Several frameworks offer best practices for the governance of information technology systems, including:

- ISO/IEC 38500, "*Governance of IT for the Organization"*
- COBIT[143] provides a reference model of thirty-seven IT processes typically found in an organization. COBIT is regarded as the world's leading IT governance and control framework.
- IGPMM- The Information Governance Process Maturity Model[144] focuses on establishing and maturing 22 processes that help identify – and improve the management of – information value, cost and risk.

ISO/IEC 38500[145], "*Governance of IT for the Organization"*, was initially adopted in 2008 as a corporate standard and then revised to apply to organizations more generally in 2015. Three major activities for an IT governing body are described[146]:

- Evaluation. The governing body evaluates the organization's overall use of IT in the context of the business environment, directs management to perform a range of tasks relating to use of IT, and continues to monitor the use of IT in the context of business and marketplace evolution.
- Assessment. Business and IT units collaboratively develop assessment proposals and plans for business strategy, investment, operations, and policy for the IT-enabled business; and
- Implementation. The governing body evaluates the proposed assessment proposals and plans and, where appropriate, directs that they should be adopted and implemented; the governing body then monitors implementation of the plans and policies as to whether they deliver required performance and conformance.

Terms such as "business environment" and "business units" can be easily transposed into the context of a digital platform enabling data exchange.

These governance activities respect the following principles:

- Responsibility. Establish appropriate responsibilities for decisions relating to the use and supply of IT;
- Strategy. Plan, supply, and use IT to best support the organization;
- Acquisition. Invest in new and ongoing use of IT;
- Performance. Ensure IT performs well with respect to business needs as required;

---

[143] Harguem, S. . (2021). A Conceptual Framework on IT Governance Impact on Organizational Performance: A Dynamic Capability Perspective. *Academic Journal of Interdisciplinary Studies*, *10*(1), 136. https://doi.org/10.36941/ajis-2021-0012

[144] Smallwood, Robert F. (2018-10-01). *Information Governance for Healthcare Professionals: A Practical Approach*. Taylor & Francis. ISBN 9781351339728.

[145] http://www.38500.org/

[146] "To Govern IT, or Not to Govern IT? " Carlos Juiz, Mark Toomey. Communications of the ACM, February 2015, Vol. 58 No. 2, Pages 58-64 10.1145/2656385

- Conformance. Ensure all aspects of decision making, use, and supply of IT conforms to formal rules; and
- Human behaviour. Ensure planning, supply, and use of IT demonstrate respect for human behaviour.

ISO/IEC 38500 differs from other IT governance frameworks (including both COBIT and IGPMM) by focussing on how decisions are made (the "responsibility" principle above) to encourage desirable behaviour in the use of IT, rather than focussing on processes, where "the best process model is often readily defeated by poor human behaviour"[147]. This approach also aligns with the broad definition of governance offered in Chapter 2, focussing on decisions, decision rights and accountability.

All three frameworks align well with the lifecycle model of governance presented in Chapter 2, highlighting different requirements at the "Formation", "Operation", "Monitoring" and "Sustainability" stages.

10.2. Cybersecurity

As noted in section 3.1, secure information systems are required by several EU laws and regulation, specifically for systems working with either personal data or data related to "critical infrastructure". The most applied international standards come from the ISO/IEC 27000-series "ISMS Family of Standards", which provides best practice recommendations on information security management—the management of information risks through information security controls—within the context of an overall Information security management system (ISMS).

- ISO/IEC 27001:2022[148] "Information security, cybersecurity and privacy protection - Information security management systems – Requirements" formally specifies a management system intended to bring information security under explicit management control.

- ISO/IEC 27701:2019[149] is a privacy extension to ISO/IEC 27001. The design goal is to enhance existing ISMSs with additional requirements to establish, implement, maintain, and continually improve a Privacy Information Management System (PIMS). The standard outlines a framework for Personally Identifiable Information (PII) Controllers and PII Processors to manage privacy controls to reduce the risk to the privacy rights of individuals.[150]

ISO/IEC 27001 outlines requirements for an ISMS capable of ensuring Information security, cybersecurity, and privacy protection, addressing:

- Formation, organized under topics of leadership, planning and support,
- Operations, and

[147] Ibid.

[148] ISO/IEC 27001:2022 - Information security, cybersecurity and privacy protection -- Information security management systems -- Requirements". International Organization for Standardization. Retrieved 13 February 2023.

[149] https://www.iso.org/standard/71670.html ISO/IEC 27701:2019 [ISO/IEC 27701:2019]

[150] "Protection of personal data: How Voluntary Standards Contribute". *AFNOR Marketing*. July 2018. Archived from the original on 2020-09-19. Retrieved 2018-07-20

- Monitoring of both performance and to enable continuous improvement.

The standard can be translated into governance and operational requirements that must be specified for the Digital Platform, as well as for the organizations involved in its operation (e.g., any organizing entity for a Data Space, the Data Intermediaries operating Data Intermediation Services, as well as any organization providing services through the Data Space which might work with data that by law can only be processed by secure information systems).

The standard details 93 information security controls in four main categories:

- Organizational Controls (37 controls). These would apply to any of the organizations listed above.
- People Controls (8). These would apply to any individuals working with data that by law can only be processed by secure information systems, not only employees of the organizations listed above, but also participants in data transactions involving data requiring security.
- Physical Controls (14). These would apply to any of the organizations listed above.vb
- Technological Controls (34). These would apply to the relevant IT systems operated by any of the organizations listed above in which data requiring security might be stored, transmitted or processed.

11. ANNEX I: Legal and Ethical Assessment Methodology

The Legal and Ethical Assessment Methodology provided by the Ethics Advisor of the GREAT project, serves as a comprehensive framework designed to systematically identify, evaluate, and address legal and ethical risks associated with a project's deliverables. Following a "by design" approach, this methodology is seamlessly integrated into the project's technical workflow, ensuring the consideration of legal and ethical aspects throughout the project's lifecycle. Its primary objectives encompass optimizing technical and business goals, ensuring compliance with relevant legal standards and ethical principles, and fostering ongoing competence-building within the research community involved.

Implemented in three key steps, the methodology begins with a preliminary meeting involving Work Package (WP) leaders, where the foundational literature and guiding legal and ethical principles are presented. The checklist analysis phase follows, employing a proactive "learning-by-doing" approach to identify potential gaps and risks across domains such as Data Privacy, Ownership, Licenses, Competition, Artificial Intelligence, and Social Media. Feedback from the Ethics Advisor on identified gaps and risks is integrated into the final deliverable, concurrently nurturing the skills necessary for crafting resilient legal and ethical solutions. These solutions address a breadth of domains and prioritize the overall impact of the deliverable while aligning with research and business goals, fostering a comprehensive legal and ethical framework.