DELIVERABLE D3.1 - DATA GOVERNANCE & METHODOLOGIES - KNOWLEDGE INITIALISATION

PROJECT ACRONYM:	DS2
PROJECT TITLE:	DataSpace, DataShare 2.0
GA NUMBER NO.	101135967
WEBSITE:	www.dataspace2.eu
DUE DATE OF DELIVERABLE:	31.12.2024
SUBMISSION DATE:	11.2.2025
LEAD BENEFICIARY:	UOS
LEAD AUTHOR:	Stefano Modafferi
REVIEWERS:	Andra Tanase; Stuart Campbell
TYPE:	R
DISSEMINATION LEVEL:	Public

This document is a deliverable of the DS2 project, which has received funding from the European Union's Horizon 2020 Programme under Grant Agreement (GA) #101135967.

DISCLAIMER

STATEMENT OF ORIGINALITY

This Deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation, or both.

DOCUMENT HISTORY

VERSION	DATE	DESCRIPTION	NAME	ORG
V0.1	30.09.2024	Toc		UoS
V0.2	01.11.2024	First draft		UoS/VTT
V0.3	30.11.2024	Second draft		UoS/VTT/INDRA/ATC
V1.0	13.12.2024	Final version for internal review		UoS
V1.1	23.12.2024	Reviewed version	A. Tanase	ATIT
V2	08.01.2024	Reviewed version	S. Campbell	ICE
V3	26.01.2025	First version addressing comments		UoS
V4	28.01.2025	Second version addressing comments		UoS
V5	06.02.2025	Submitted version		UoS

TABLE OF CONTENTS

EXECUTIVE SUMMARY

This deliverable provides a comprehensive exploration of trust and data sovereignty within the evolving European data sharing landscape, focusing on the critical challenges and innovative approaches to facilitating B2B data exchange.

Key Objectives

The deliverable investigates the complex dimensions of trust and sovereignty in data spaces, analysing the barriers to data sharing, the European regulatory framework, and its implications for cross-sector data sharing, and proposing a model and a methodology for data lifecycle management, finally linking these aspects to DS2 modules.

The human perspective

A targeted survey was developed to capture explicit and implicit barriers to data sharing, engaging stakeholders with decision-making influence in B2B contexts. The survey's structure is included in the annex and the data analysis (to be presented in D2.2) will provide critical insights.

The regulatory landscape

A detailed analysis of European Commission regulations is provided, and it discusses the multifaceted legal environment governing data sharing. The deliverable maps the regulatory requirements and their potential impact on inter-organizational data exchange, providing insights into the evolving legal frameworks that shape data sovereignty.

Trust and Risk Management

D3.1 introduces a robust risk management strategy designed to build trust among data sharing participants. By conceptualizing trust in B2B data sharing and presenting innovative governance mechanisms, the research offers a nuanced framework for addressing organizational hesitations.

The product-based data lifecycle management

D3.1 exploits an innovative approach to data lifecycle management centred on the concept of a data product. This approach transforms data into a managed, valuable asset, emphasizing reusability, governance, and strategic value creation

Relationship with Key Modules

D3.1 also links three essential modules—DSM (Sovereignty Decision Support), DRM (Digital Rights Management), and PAE (Policy Enforcement Module)—to the analysis provided. These tools are designed to operationalize the concepts discussed throughout the report, facilitating effective governance and management of data sharing processes. The DSM module supports decision-making regarding sovereignty issues, while the DRM tool leverages blockchain technology for secure digital asset management. The PAE module ensures compliance through policy enforcement and robust action traceability, all integral to enhancing trust in B2B interactions.

This deliverable represents a critical step in understanding and addressing the complex challenges of data sharing, providing a foundation for more effective, secure, and collaborative data ecosystems.

1 INTRODUCTION

The Data Space | Data Share (DS2) project (https://www.dataspace2.eu/) draws researchers and practitioners from diverse disciplines to guarantee that the complex lifecycles of inter-sector data sharing, aggregation, and provenance will take place in a human-centric and trusted way, with common structures, exportability and insight, whilst protecting the sovereign rights of data owners and complying with European data regulations. As stated in the DS2 D2.1 report on problem space definition[1], the project follows the meaning of the term data space given in the Data Spaces Support Centre (DSSC) Glossary[2]:

> *"A distributed system defined by a governance framework that enables secure and trustworthy data transactions between participants while supporting trust and data sovereignty. A data space is implemented by one or more infrastructures and enables one or more use cases" (bold emphasis in original text) (Poikola et al., 2023).*

Further confirmed by the definition from the Agencia Española de Protección de Datos (AEPD, 2023) — the Spanish Data Protection Authority — as

> *"a federated and open infrastructure to allow sovereign access to data, based on governance, policies, rules and standards that define a framework of trust for all parties involved."*

DS2 provides a modular software infrastructure to connect data sources (data spaces/data silos/data lakes) together for the purpose of cross-sector data sharing. Once connected, data consumers and data providers will be able to structure and execute efficient complex data lifecycles that respect the technical and governance related requirements of the participating data sources. It will do this via an IDT (Inter-sector Data Space Toolkit) which is deployed in front of each data source/space and network connected to any other IDT-enabled data source. The IDT is composed of a Broker which manages the fail-safe network operation with no central point of control. Plugged into this is a set of modules for the execution of complex data lifecycles — e.g., filtering, labelling, both automated and accommodating for where human-in-the loop is required.

DS2 will pilot and evaluate its technology using three well-defined, inter-sector use cases — i.e., City Scape, Green Deal, Precision Agriculture — and one cross-inter-sector use case[3]. The DS2 solution enhances and accelerates the shift towards the data-agile economy by addressing the challenges, pain-points, and requirements with respect to the execution of complex data lifecycle. Data consumers and data providers will now be able to orchestrate, manage, and securely execute complex data lifecycles to realise cross-sectorial data driven applications.

The successful implementation of data sharing initiatives relies heavily on understanding the complex human factors that hinder effective data exchange. Many organizations view data sharing primarily through a lens of risk, associating it with regulatory challenges and potential repercussions. This risk-averse mindset often overshadows the significant benefits of collaboration, leading to missed opportunities for innovation and value creation. By gaining insights into user attitudes and a better understanding of the regulatory framework, organizations can develop governance frameworks that address these concerns, clarify data sharing processes, and demonstrate the tangible value of collaboration. This understanding is not just an academic exercise; it is

[1] https://www.dataspace2.eu/results/deliverables
[2] https://dssc.eu/space/Glossary/
[3] https://www.dataspace2.eu/usecases

essential for fostering open and trustworthy data ecosystems that can drive meaningful societal and economic progress.

1.1 Objectives and structure of this report

WP3 goal is to address critical aspects of data sovereignty and data sharing in the evolving landscape of data spaces. This paragraph explains the link between project tasks, objectives and key exploitable results, and the sections in the document.

This first version of the WP3 deliverable supports the understanding of the general landscape in data sharing both form a human attitude and regulatory perspective, and then introduces the relevant methodologies (data product lifecycle management and risk assessment for trust building) to support it, and finally moves to present the three main DS2 tools relevant for data sharing governance and methodologies (namely DSM, DRM, PAE).

Activities in T3.4 (Motivations, Barriers & Strategies for Data Sharing & Access) aligns with O3.4, determining motivations and barriers for data sharing. The report starts with this section (section 2) to acknowledge the centrality of the user, user decision and user attitude to successfully implement the data sharing. Sharing data is primarily a willing decision by individuals and organizations. Without analysing them, it is almost impossible to understand how to support a better, wider, and more effective data sharing. The section discussion contributes to providing DS2 Enhanced Knowledge of barriers and motivations for data sharing (KER 3.4).

Activities in T3.1 (Complex Data Lifecycle Modelling) aligns with O3.1, which focuses on modelling data sovereignty throughout complex lifecycles and in the report is primarily addressed in Section 3, which explores the meanings and interpretations of data sovereignty in data spaces. The theoretical bases discussed in this section are complemented by the legislation landscape discussion of Section 4 and then make operational in Section 5.

Activities in T3.2 (GDPR, Ethics & Policy Compliance Data regulation modelling) corresponds to O3.2, providing support for regulatory compliance. This is covered in Section 4, which examines the EU regulatory landscape concerning data and provide the backbone for the DS2 Regulatory Guidance for Practitioners (KER3.2).

The transitioning from understanding the context to provide actual support to data sharing requires an operational definition of data lifecycle management which Is addressed in Section 5, and it is based on the concept of data product.

Moving to the software-based support, three tools are mainly related to the aspects defined in WP3 and they are discussed and presented in Section 6: the DSM module (providing the sovereignty decision support system - KER 3.1), the DRM tool to provide a blockchain based house clearing (DRM – KER 3.3) the policy enforcement module (PAE) actually developed in T4.1. The impact of the finding of WP3 on the other tools will be addressed in D3.2.

Other module might be affected by the findings of WP3 and this will be investigated and addressed in the next version of this deliverable D2.2 which will be composed of both a reporting and a software part.

Annex A is the bibliography, Annex B presents the document history and Annex C the survey structure.

1.2 Glossary and Abbreviations

A definition of common terms related to DS2 as well as a list of abbreviations, is available at https://www.dataspace2.eu/results/glossary

2 EXPLORING USERS' ATTITUDES TOWARDS B2B DATA ACCESS AND USE IN DATA SPACES

A key objective for WP3 is to examine the motivations, barriers and strategies for data sharing and access in data spaces (T3.4). To do this, this deliverable utilises methods such as desk research, crowdsourced surveys, focus groups, semi-structured interviews to consult with different stakeholders and determine the level of understanding of the implications of sharing data, and the attitudes towards the acceptability of sharing data, and finally to determine the differences between attitudes and actual actions that people have taken regarding sharing data. This report provides a brief overview of some of the existing challenges with business-to-business data sharing and re-usage, and the motivation and opportunities for data spaces in this context (section 2.1). It then describes the overall design considerations for a survey concerning users' inter-organisational data attitudes and behaviours together with the accompanying dissemination strategy (section 2.2).

2.1 Fostering greater business-to-business (B2B) data sharing and re-use

Business to business (B2B) data sharing and re-use is broadly described by the European Commission et al. (2018) as "making data available to or accessing data from other organizations for business purposes". A business may be willing to share data with other businesses for various reasons, such as for "remuneration" in terms of "the exchange data for data of equal perceived value", data being "exchanged for services or access to an improved service based on shared data", and data being "exchanged for money" (Jussen et al., 2024). In addition to these motivations, businesses often engage in B2B data sharing to enhance operational efficiencies, drive innovation, and improve customer experiences, organizations may seek to leverage shared data to develop new products or services, optimize existing offerings, and gain competitive advantages in the marketplace (Martens *et al.*, 2020). Moreover, organizations might share data to comply with regulatory requirements or industry standards that encourage transparency and collaboration. The potential for cost savings through shared resources and collective problem-solving also serves as a compelling incentive (Martens *et al.*, 2020). Further, a business may also be willing to donate data to others, such as for "research purposes without expecting a direct service in return" (Jussen et al., 2024).

2.1.1 Concerns with B2B data sharing and re-use

The European Strategy for Data (European Commission, 2020) suggests that B2B data sharing and re-use "has not taken off at sufficient scale" owing to "a lack of economic incentives [...], lack of trust between economic operators that the data will be used in line with contractual agreements, imbalances in negotiating power, the fear of misappropriation of the data by third parties, and a lack of legal clarity on who can do what with the data [...]". Aligned with this view, other commentators have also highlighted concerns over "leakage of competitive knowledge and intellectual property" (Fruhwirth, Pammer-Schindler and Thalmann, 2024) as well as "competition, privacy, and reputational risks" (Bernal, 2024). In the words of (Agahari, Ofe and de Reuver, 2022): "Firms are often reluctant to share data because of mistrust, concerns over control, and other risks". Major barriers are also around the lack of awareness of good practices and the details of mechanisms of data sharing (Fassnacht, M. *et al.* ,2023)

2.1.2 Motivation and opportunities for data spaces

Facilitating greater sharing and re-use of privately held data between organisations is therefore presented as one of the principal aims of the European Strategy for Data (European Commission, 2020). One approach to fostering increased B2B data sharing and re-use centres on the creation of common European data spaces, which strive to address "legal and technical barriers to data sharing across organisations" as well as "issues of trust" (European Commission, 2020). A data space is described by the Agencia Española de Protección de Datos (AEPD, 2023) — the Spanish Data Protection Authority — as "a federated and open infrastructure to allow sovereign access to data, based on governance, policies, rules and standards that define a framework of trust for all parties involved." Given that the multiple ways in which the distribution of function (e.g., Boniface et al., 2020) can be configured within a dataspace or a federation of dataspaces (AEPD, 2023), the term data space should be understood as an "umbrella term corresponding to any ecosystem of data models, datasets, ontologies, data sharing contracts, and specialized management services [...] together with soft competencies around it [...]" (Scerri *et al.*, 2022). As examples of efforts towards data space standardisation, consider the Gaia-X[4] and International Data Space Association (IDSA)[5] approaches (e.g., Braud et al., 2021).

Some key opportunities to "generate value" from common European data spaces for "business", "citizens", "science" and "government and public bodies" have been highlighted by the Big Value Data Association (BVDA) (Scerri et al., 2022). In terms of value for "SMEs and large industries", four key opportunities are identified, which are:

- "Open data marketplaces that level the playing field for industrial data sharing"
- "Increased availability of vast and heterogeneous data ecosystems for AI"
- "Innovative data-driven business models enabled by new value ecosystems", and
- "Opportunities to tap into 'safe' personal data" (Scerri et al., 2022)

More recently, in a study conducted by (Hutterer & Krumay 2024) that involved 28 qualitative interviews with selected experts, twelve "drivers influencing data space adoption" were identified — these are: "Controllable complexity", "Cost clarity", "Data sovereignty", "Ecosystem governance", "Ecosystem readiness", "Interoperability", "Mature technology", "Regulatory certainty", "Security", "Technology competence", "Transparency" and "Trust".

2.1.3 User Inter-organisational Data Sharing Attitudes and Behaviours Survey

This section describes the overall design considerations of the "User Inter-organisational Data Sharing Attitudes and Behaviours Survey" and the accompanying dissemination strategy. The aim of the survey is to provide an understanding of the existing attitudes of key stakeholders, with a particular focus on stakeholders within the DS2 Use Case domains of Cityscape, Green Deal and Precision Agriculture, to the sharing of data between collaborative business partners, and to gain some knowledge of current inter-organisational data sharing practices. This will be used to inform further WP3 and DS2 work by providing insight into the barriers to the adoption of data sharing practices and how they might be overcome through considered design and development of legal and governance structures and technologies.

[4] https://gaia-x.eu/ (Accessed 25 October 2024).
[5] https://internationaldataspaces.org/ (Accessed 25 October 2024).

2.1.3.1 Survey Design

The survey was developed as a mixed method, fully anonymised, online survey using the University of Southampton approved and secure instance of the Qualtrics platform. Ethical approval for the survey was awarded under ERGO ID 100306. All relevant participant information, opt in and consent was provided at the start of the survey. There were no specific gender or other Equality Diversion Inclusion (EDI)[6] considerations as no personal or demographic information was requested.

The survey targets senior managers from public and private European and UK organisations, such as CTO/CIOs, Data Officers/Managers/Analysts…etc, with some degree of responsibility for data management and sharing. It consists of 36 mainly multiple-choice and yes/no questions, with 3 open questions (2 of which are optional, nested questions). Several drafts of the survey were reviewed and improved by members of the WP3 team over several weeks, resulting in the fine-tuning of many questions and answer options, and the addition or removal of entire questions. A final draft version was then trialled with 6 individuals for errors and time-to-complete. Further changes were made and the time-to-complete reduced to under 15 minutes, at which point the survey was published.

The survey was divided into 4 main sections:

1. About You
2. Attitudes to Inter-organisational Data Sharing
3. Your organisation's current practices
4. Data Sovereignty

Each section was designed to provide specific information to support the survey's objectives.

Section 1: About You was designed to serve firstly as a filter to remove non-qualifying respondents – those with no knowledge, responsibility or decision-making power for inter-organisational data sharing and respondents from organisations not currently engaged in data sharing. For those respondents qualifying for the survey the remaining part of the section captures some (anonymous) information concerning job role, organisational domain, data sharing role, and understanding of Data Sharing Agreements. This will enable an analysis of attitudes and behaviours for the whole group and also for sub-groups based on the Section 1 variables (e.g. domain).

Section 2: Attitudes to Inter-organisational Data Sharing was designed to explore the attitudes and willingness to engage in inter-organisational data sharing and the challenges in doing so, as well as familiarity with and attitudes to existing legal frameworks for data sharing. Questions were developed from a review of the existing recent literature in this domain (e.g. (Fassnacht *et al.*, 2023; Jussen, Schweihoff and Möller, 2023; Jussen *et al.*, 2024), which informed a division of challenges into separate questions focusing on strategic, operational, network/partnership, and technical barriers to adoption. This will allow a nuanced insight into these barriers thereby helping to directly inform WP3/DS2 design considerations.

Section 3: Your organisation's current practices was designed to establish an understanding of what exactly is going on in this domain at the moment. It explores risk assessment practices; data protection priorities; existing infrastructure and processes; data value, governance, & training; and data types, formats, licences,

[6] https://en.wikipedia.org/wiki/Diversity,_equity,_and_inclusion

and restrictions. This will provide a detailed picture of the existing inter-organisational data sharing landscape which will further support WP3/DS2 design and development activities by highlighting gaps, patterns, relationships, and common behaviours.

Section 4: Data Sovereignty was included as this notion is subject to considerable confusion, even within the existing literature (e.g. von Scherenberg et. al., 2024; Hummel et. al., 2021), with some stakeholders viewing it as mainly referring to jurisdictions, laws and regulations and others primarily relating it to data ownership, rights and control. Given this, it was decided that a small separate section on Data Sovereignty would be of value in providing an understanding of whether this polarity is reflected in reality, or whether there is a dominant understanding of the term among stakeholders.

The survey will provide results that can be descriptively analysed, and if necessary or useful, can also be tested for significance, with findings able to support, clarify or refute the existing literature as well as help guide and inform DS2 legal, governance and technical outputs, including the development of the T3.1 Knowledge Base.

2.1.3.2 Survey Dissemination Strategy

The survey dissemination strategy consists of 3 activity strands:

- Leveraging DS2 consortium partner's networks
- Social media – focusing on LinkedIn
- European Data Space Stakeholder Organisations

Market research companies (e.g. userinterviews; lyssna) were explored as a potential route for dissemination, and virtual meetings were held to discuss capabilities and costs. However, it was decided that the reach into the target audience was limited - both geographically and in terms of fit – and the costs were prohibitively high for that sort of return. As a result, this dissemination channel was discarded and instead the following approaches were made:

- Strand 1 involves providing all DS2 partners with an email link and description of the survey and a request to forward the same across the relevant members of their business and government networks. As stakeholders in the domain with wide networks of other domain stakeholders this Strand will ensure dissemination to qualifying individuals / organisations.

- Strand 2 involves developing a LinkedIn post on the DS2 account providing the survey link and purpose and then alerting all DS2 partners to like and repost across their own LinkedIn networks. This will enable a wide reach, but with only a moderate focus on the target audience.

- Strand 3 involves dissemination of the survey link and purpose to EU and UK umbrella organisations such as the BDVA (Big Data Value Association), CEDPO (Confederation of European Data Protection Organisations), IDSA (International Data Spaces Association), GAIA-X, FIWARE, and other stakeholder organisations with a request to notify members and networks of the survey. As a considerable amount of current inter-organisational data sharing occurs between public sector bodies, this Strand also includes engagement with public organisations such as the EDIB (European Data Innovation Board), ADR UK (Administrative Data Research UK), and the CDDO (Central Digital Data Office). This will facilitate dissemination across target networks and extend reach into public sector organisations.

The target number of responses is 50-100. Currently, Strand 1 was initiated in M11, Strands 2 and 3 follow in M12. The full survey can be found in Annex C. Survey results and findings will be made available to the

consortium through plenary presentations and to the wider stakeholder community through journal or proceedings publications.

3 UNDERSTANDING MEANINGS AND INTERPRETATIONS OF DATA SOVEREIGNTY IN THE DATA SPACES CONTEXT

Support for data sovereignty is presented as an essential feature of data spaces in the data space definition provided by DSSC definition of a data space (Poikola et al., 2023). Yet, various meanings and interpretations are given to the concept of data sovereignty in the data spaces context (e.g., Ryan et al., 2024; von Scherenberg et al., 2024) and more broadly (e.g., Hummel et al., 2019). A key action for WP3 therefore is to scope and define different interpretations of data sovereignty from the various perspectives of the key stakeholders (T3.1). To do this, it will conduct desk research as well as consult with key stakeholders to identify the key principles of data sovereignty and their different priorities and concerns in the context of data spaces. This report provides an initial overview of data sovereignty conceptualisation in the data spaces context (section 3.1). A summary the key points about the meanings and interpretations of data sovereignty in data spaces from this literature review will follow, which can help to inform our methodology for consulting with key stakeholders (section 3.2). The aim of this consultation is to provide further insight into the meaning of data sovereignty from different stakeholder perspectives, and to identify different events in complex data lifecycles and how these events affect sovereignty in the context of data spaces.

3.1 Data sovereignty conceptualisation

The notion of data sovereignty is widely used in relation to data spaces (note: some example definitions are provided in Section 3.1.1.1). By way of illustration, according to Bacco et al. (2024): “From a policy perspective, a data space represents a distributed system governed by a framework that enables secure and trustworthy data transactions while maintaining data sovereignty”. Another example being that “sovereign” features as one of the Gaia-X ten core values (Bonfiglio, 2021). However, it is important to recognise that while the notion of data sovereignty is presented as a fundamental aspect of data space theory and practice (e.g., Agencia Española de Protección de Datos [AEPD], 2023a; Hutterer & Krumay, 2024; Poikola et al., 2023), there is no legal definition given for this concept in EU data law (Ryan et al., 2024).

Various meanings and interpretations are given to the notion of data sovereignty across different disciplines, domains, and contexts both more generally (e.g., Hummel et al., 2019) and in specific relation to data spaces (e.g., Ryan et al., 2024; von Scherenberg et al., 2024). For instance, indigenous data sovereignty “draws on the United Nations Declaration on the Rights of Indigenous Peoples (UNDRIP), which reaffirms the rights of Indigenous Peoples to control data about their peoples, lands, and resources” (Carroll et al., 2021). It should be noted that data sovereignty also can be used in reference to data localisation — e.g., defined as “a mandatory legal or administrative requirement directly or indirectly stipulating that data be stored or processed, exclusively or non-exclusively, within a specified jurisdiction” (Svantesson, 2020).

3.1.1 A need for a shared understanding of data sovereignty in the data spaces context

It is essential that the “conceptual challenges” (Ryan et al., 2024) around the use of the term data sovereignty are acknowledged (von Scherenberg et al., 2024), and it is recognised that the term data sovereignty may have distinct meanings in different contexts. This is important given that a lack of shared understanding about what data sovereignty means between different data space actors for a specified context may give rise to issues around e.g., misconceptions related to the plans and actions necessary to support data sovereignty in practice

(Ryan et al., 2024) as well as misalignment between expectations for B2B data sharing and re-use within a particular data space.

Reviews into the various meanings and interpretations of the notion of data sovereignty have been undertaken. One example being the literature review provided by Hummel et al. (2019), which found from the "candidate meanings" of which they identified that definitions of data sovereignty "tend to relate in some way to meaningful control, ownership, and other claims in data". By using the term "meaningful control", the intention is that the type of control is "e.g. one with epistemic and social presuppositions, such as awareness of the potential scope of data processing applications and entitlements to control resulting from recognition of one's fundamental rights" (Hummel et al., 2019). On a similar note, Abbas et al. (2024) caution against the "conceptual reductionism (i.e., an oversimplification of a complex phenomenon)" whereby narrow control-centric interpretations — i.e., the focus on control as "the capability to influence and direct information flows" — overlook other crucial aspects of data sovereignty that extend beyond an organisation's ability to control the data they share with other organisations. Such other important aspects related to data access and use include:

- "ownership" — i.e., "data property rights, indicating the privileges over data resources"; "privacy" — i.e., "encapsulates the protection of personal data";
- "security" — i.e., "focuses on preventing potential threats and risk mitigation concerning data";
- "responsibility" — i.e., "delineates roles and expectations"; and,
- "compliance" — i.e., "adherence to relevant legal and regulatory frameworks" (All quotes in bullet point list from: Abbas et al., 2024).

As an additional example, for effective data sovereignty, data spaces also need to be sustainable— this has been referred to as "data gravity" by Marino et al. (2023) — i.e., "the ability of data to attract applications, services, and other data".

Further, Ryan et al. (2024) have reviewed the main characteristics of data sovereignty as specifically applied to data spaces from selected literature which are:

- "The values that underpin data sovereignty" — in particular "data ownership", "transparency", "privacy" and "control/power/authority" and other associated values
- "The data processes involved" — which enable "control over the access, use, sharing, and storage of one's data"
- "The data sovereign agent" — such as, "individuals, organizations, and states" (All quotes in bullet point list: Ryan et al., 2024)

3.1.1.1 Some example definitions

For purposes of illustration, here are some example definitions of data sovereignty as applied to data spaces given in the literature (for further examples of definitions and etymology of the term sovereignty, see e.g., Hummel et al., 2019; Cheung, 2024; Ryan et al., 2024; von Scherenberg et al., 2024):

Source	Some examples of data sovereignty definitions in data space related literature

Agencia Española de Protección de Datos (AEPD). (2023a).	"a concept not defined in the European standard and generally interpreted as the idea that the place where data is collected determines the regulation and governance that applies to it, and also the ability of governments and companies to use of users' and companies' digital data". [General definition]
AEPD (2023b).	"That control, and the trust the stakeholders need in the data-access sharing economy, is called "data sovereignty"." [General definition]
Bacco et al. (2024).	In this context, "sovereign exchange of data" can be seen as "being able to self-determine who, how, when, and at what price others may use data across the value chain" [Sovereign exchange of data]
Data Spaces Support Centre (DSSC, 2023)	"The ability of individuals, organisations, and governments to have control over their data and exercise their rights on the data, including its collection, storage, sharing, and use by others. [/] Explanatory text: Data sovereignty is a central concept in the European data strategy and recent European laws and regulations are expanding upon these rights and controls. EU law applies to data collected in the EU and/or about data subjects in the EU" [General definition]
Duisberg (2022).	"The claim of data sovereignty is inherently linked to putting the legal instruments and tools in the hands of each participant in the ecosystem, allowing freedom of contract as well as ensuring that exercising data exchange and consorted data usage in the data economy is in compliance with general and specific regulations, ranging from anti-trust to GDPR and cyber-security regulations as well as sector specific regulations" [Contract law perspective]
Falcão et al. (2023) [Note: Also cited by Sullivan et al., 2024]	"We define data sovereignty (DS) in the agriculture domain based on three pillars: data portability (the possibility to move data from one system to another), data usage only with consent, and transparency about what happens with the data" [Domain-specific definition: Agricultural data spaces]
Bonfiglio (2021); Gaia-X (2023)	"Sovereignty is the ability to exercise self-determination. It can translate into several meanings- political, economic, digital, and technical. Gaia-X does not provide any political or economic interpretation of sovereignty, but instead provides a framework to configure sovereignty from a digital and technical perspective."
International Data Spaces Association (IDSA, 2024)	"What is data sovereignty? Today, organizations of all types and sizes collect and store huge amounts of every kind of data. IDSA enables you to self-determine how and when others may use it across the value chain. We call this data sovereignty." [General definition]
Jarke et al. (2019)[7]	"Data sovereignty refers to the self-determination of individuals and organizations regarding the use of their data. […] data sovereignty aims at enabling "data richness" by clearly negotiated and strictly monitored data usage agreements." [General definition]
Otto (2022)	"Data sovereignty refers to the capability of a legal entity or natural person to determine and execute usage rights when it comes to their data."
Pettenpohl et al. (2022)	"Data sovereignty is a fundamental aspect of the International Data Spaces (IDS). It can be defined as a natural person's or corporate entity's capability of being entirely self-determined regarding its data. This means that a data owner can define usage restriction to their data, before sharing it with data consumers. Data consumers must accept the usage restrictions." [General definition]

[7] This definition is referred to by e.g., Hutterer & Krumay (2024).

Ryan et al. (2024)	"Data sovereignty in data spaces is control by an individual, organization, or state over the access, use, storage, and sharing of their data" [Working definition]

Table 1 Some examples of data sovereignty definitions in data space related literature

3.1.1.2 Different perspectives of data sovereignty

The notion of data sovereignty can be viewed from multiple angles — e.g., from those of different sovereign agents. For instance, from the view of individuals, consider the notion of "data self-sovereignty" defined by Cheung (2024) as "the empowerment of the self to have effective and meaningful control over one's data". This notion may also be referred to as personal data sovereignty (e.g., Micheli et al., 2020; Carmichael et al., 2024). Further to this: "The concept of data sovereignty encompasses entitlements of the individual to connect and share information with others. It thus demands not merely constrainable but controllable data flows" (Hummel et al., 2018). Moreover, from an organisational perspective, the term "organizational data sovereignty" is used as a way to describe "granting organizations control over their data" (Opriel *et al.*, 2024) — as well as the promise of "fair distribution of benefits and control over one's data resources while emphasizing data creation value [...] and increased economic efficiency for data-driven businesses" (Ryan et al., 2024). Increasing organisational data sovereignty is a key focus for the DS2 project and is promoted as one way in which barriers to inter-organisational data access and usage can be overcome — especially those related to concerns over "disclosing valuable information that could endanger a stakeholder's competitive advantage" (Heeß et al., 2024).

An additional concept of self-determination is introduced by the definitions of data sovereignty given by e.g., Bonfiglio (2021), Jarke et al. (2019) — as shown in the table above. A definition for the concept of "digital self-determination" is given by Verhulst (2023) as "the principle of respecting, embedding, and enforcing people's and peoples' agency, rights, interests, preferences, and expectations throughout the digital data life cycle in a mutually beneficial manner [...] for all parties [...] involved" (note: footnotes from original text omitted). A key point to highlight here is that data sovereignty principles should not be viewed as "absolute", but rather "shared" between multiple parties involved in data spaces (Ryan et al., 2024). On a related note, the principles of data sovereignty can be seen as a "spectrum" (IDSA, 2024) where an appropriate balance needs to be determined between significant properties, such as "between keeping data safe and sharing it to gain added value" (IDSA, 2024) with the wider data access and usage ecosystem.

3.1.2 Putting data sovereignty principles into practice

A further issue concerning meanings and interpretations of data sovereignty is that definitions often lack sufficient information about what organisational and technical mechanisms should be used to support the exercise and implementation of data sovereignty in practice (Ryan et al., 2024; von Scherenberg et al., 2024). Further, appropriate assurances need to be given to organisations about how and to what extent data sovereignty can be operationalised by a specified data space in practice (Heeß et al., 2024). In view of this, it is worth mentioning that the "key features of a common European data space" outlined in the Commission Staff Working Document on Common European Data Spaces (European Commission, 2022) do not only necessitate "secure and privacy-preserving infrastructure" but "also clear and trustworthy data governance mechanisms" — together with other requirements e.g., related to respecting "European rules and values". Further, in relation to "operationalizing digital self-determination", Verhulst (2023) provides a "four-pronged framework" centred on both "human or human initiated processes" and "technology", which comprises:

- "processes" — e.g., "data assemblies"
- "people and organizations" — e.g., "data stewards", "data intermediaries"

- "policies" — e.g., "charter", "social license", "code of conduct" and
- "products and technologies" — e.g., "a trusted data space"

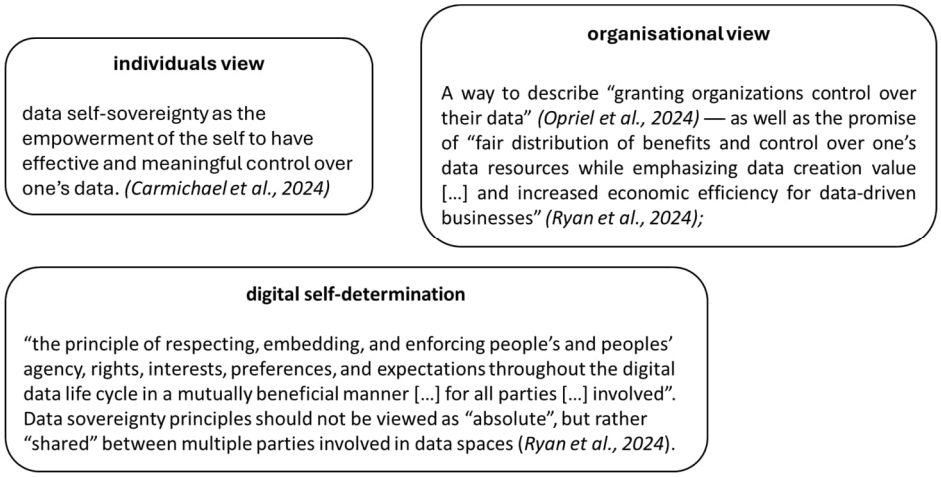

Figure 1 Different Perspectives on Sovereignty

Some examples of such mechanisms for supporting data sovereignty in practice include "data usage control policies" (Otto, 2022), "ICT security methods and tools" (Jarke et al., 2019) — e.g., "confidential computing and homomorphic encryption" (Hutterer & Krumay, 2024), "implementation of policies" (Hutterer & Krumay, 2024), measures for "data portability" (Jarke et al., 2019), "privacy-preserving approaches, and privacy-enhancing technologies" (Hutterer & Krumay, 2024). For further illustration, the approach taken by the International Data Spaces Association (IDSA) in creating their "Reference Architecture Model" should also be considered (Duisberg, 2022; Pettenpohl et al., 2022). In this Model, the Connector is presented as the "main technical component" of this model, which aims to ensure that "participants maintain sovereignty over the data" by hosting various "system services" depending on the "configuration", such as "a secure bidirectional communication, enforcement of usage policies upon exchanged content, system monitoring, and logging of content transactions for clearing purposes" (Pettenpohl et al., 2022). From a contract law perspective, according to Duisberg (2022), a principal aim of the IDSA approach to data spaces is that of "usage control" whereby the terms and condition concerning data access and use set out in specified data licensing agreements are enforced through such technical usage control mechanisms. The intention being that such "usage control will have a stronger effect than the traditional licensing models" (Duisberg, 2022).

It should also be highlighted that studies concerned with exploring the requirements for and challenges with implementing data sovereignty in practice have been carried out. One example being the study conducted by Hellmeier et al. (2023), where semi-structured interviews were carried out with eleven industry experts on this topic. In terms of "requirements for data sharing", these were divided into two categories:

- "organization requirements" related to "law, legal, restrictions and guidelines", "confidentiality", and "data/service classification"; and
- "technical requirements" focused on "policy enforcement", "data security and integrity", "data visibility and offering", and "transparency" (Hellmeier et al., 2023).

Further, in respect of challenges, these were split into three groups:

- "organizational challenges" — related to e.g., "staffing", "communication" etc;
- "technical challenges" — about e.g., "data processing lifecycle", "identity management" etc; and

- "personal and emotional challenges" — with reference to "trust" and "comfort" (Hellmeier et al., 2023). Also consider a study conducted by Hutterer & Krumay (2024) that involved 28 qualitative interviews with selected experts, where data sovereignty was identified as one of twelve "drivers influencing data space adoption".

3.1.3 Examples of some existing data sovereignty conceptual models and frameworks

The report now examine two instances of existing conceptual models and frameworks from the literature — as presented by Abbas et al. (2024) and von Scherenberg et al. (2024) — to provide examples of the ways in which key aspects of data sovereignty have been identified and represented from an information systems perspective.

Abbas et al. (2024) explore the notion of data sovereignty, in the context of organisational data access and use, through the lens of "Social Contract Theory (SCT)" where the "spatial, temporal, and substantive aspects that shape social contracts" are examined. Some key elements of this model include:

- "Control-based provision" — "facilitates horizontal interactions among actors such as data subjects and providers to ensure sovereignty"
- "Defence-based provision" — i.e., "security and compliance mechanisms"; and
- "Foundational rights that need protection" — "ownership" and "privacy"

Abbas et al. (2024) also highlight how different "contextual conditions" — e.g., "data type", "business data sharing setting", "organizational size" can influence different aspects of data sovereignty. One example being how "variations" in data format may "raise technical challenges for provisioning *control* mechanisms" (Abbas et al., 2024).

From this study, it is highlighted that "data control" mechanisms (e.g., "ongoing monitoring post-transaction" etc.), "security" mechanisms (e.g., "encryption", "watermarking" etc.), and "compliance" mechanisms (e.g., for "consistent alignment with meta-platform technical specifications", fulfilling "contractual obligations" etc.) are required for operationalising data sovereignty (Abbas et al., 2024). Further, the roles and responsibilities of data access and usage actors need to be well-defined (Abbas et al., 2024).

Further to Abbas, as an example of one approach to help provide a "consistent understanding" of the data sovereignty concept, von Scherenberg et al. (2024) have proposed a "data sovereignty conceptual model", which aims to help determine fifteen "core aspects"[8] required for "technical implementation" of data sovereignty from an information systems perspective. In brief, the concept of data sovereignty is conceived "an instrument to keep control over an actor's data asset" in this model — where access and use to a specified data asset is controlled via a "contractual agreement" negotiated by the "data provider" and "data consumer", which is in turn enforced through the "data infrastructure" (von Scherenberg et al., 2024). This conceptual

[8] These core aspects being: "data asset", "data provider", "data consumer", "contractual agreement", "data value chain and data lifecycle activities", "data infrastructure", "trust", "relations", "data provider and data consumer require trust", "data infrastructure ensures trust", "data provider and data consumer negotiate contract agreements", "data infrastructure supports management of contractual agreement", "Contractual agreement specifies use conditions of data asset", "Data provider and data consumer perform data value chain and data lifecycle activities", and "Data value chain and data lifecycle activities modify data asset" (von Scherenberg et al., 2024). See original text for more details.

model is “grounded in agency theory”, and “through the lens of this theory, data sovereignty can be implemented as an instrument with the central objective of establishing more trust” (von Scherenberg et al., 2024).

3.2 Summary: key considerations

From this literature review, some key points about the meanings and interpretations of data sovereignty in data spaces are highlighted:

- While the notion of data sovereignty is presented as a fundamental aspect of data space theory and practice (e.g., AEPD, 2023a; Hutterer and Krumay, 2024; Poikola et al., 2023), various meanings and interpretations are given to the notion of data sovereignty across different disciplines, domains and contexts both more generally (e.g., (Hummel *et al.*, 2018)) and in specific relation to data spaces (e.g., Ryan et al., 2024; von Scherenberg, Hellmeier and Otto, 2024). Further, there is no legal definition of data sovereignty given in EU data law (Ryan, Gürtler and Bogucki, 2024).

- However, there are existing publications reviewing the literature related to the etymology of the term sovereignty, and meanings and interpretations of the concept more broadly and in the data spaces context (e.g., Hummel et al., 2019; Cheung, 2024; Ryan et al., 2024; von Scherenberg et al., 2024).

- Broadly speaking, the concept of data sovereignty when used in relation to data spaces can be viewed as actors (e.g., organisations, individuals) — also referred to as sovereign agents (Ryan et al., 2024) — having some meaningful control (Hummel et al., 2019) over the processing of shared data across its lifecycle (e.g., collection, storage, access, use etc.) where they would be willing and able to make such data available. Further, in this context, data sovereignty is where actors are allowed to exercise their rights over the data that has been shared (Poikola et al., 2023) such as, contractual rights and the rights of individuals under the General Data Protection Regulation (GDPR). Further, data sovereignty appears to be strongly associated with the notion of self-determination over data (e.g., Bonfiglio, 2021; Jarke et al., 2019; Verhulst, 2023).

- The concept of data sovereignty extends beyond control and is underpinned by other important values, such as those relating to compliance, data protection, privacy, ownership, responsibility, security, and sustainability (e.g., Abbas et al., 2024; Marino et al., 2023; Ryan et al., 2024).

- Given the various meanings and interpretations of data sovereignty, it is important to foster a shared understanding of the concept and what it means for different actors in a specified data space — e.g., for organisations in terms of organisational data sovereignty (e.g., Opriel et al., 2024, Ryan et al., 2024), for individuals in respect of data self-sovereignty or personal data sovereignty (e.g., (Micheli *et al.*, 2020; Carmichael, Hall and Boniface, 2024; Cheung, 2024). A lack of shared understanding can be problematic — for instance, potentially giving rise to misalignment of expectations for data access and use, and misconceptions over the plans and actions necessary to support data sovereignty in practice (e.g., (Abbas *et al.*, 2024; Ryan et al., 2024).

- Data sovereignty definitions often lack sufficient information about what organisational and technical mechanisms should be used to support the exercise and implementation of data sovereignty in practice (Ryan et al., 2024; von Scherenberg et al., 2024). Consideration needs to be given therefore to the types of mechanisms to be deployed (e.g., Duisberg, 2022; Hellmeier et al., 2023; Hutterer & Krumay, 2024; Jarke, Otto and Ram, 2019; Otto, 2022; Pettenpohl, Spiekermann and Both, 2022; Verhulst, 2023), the

challenges in putting data sovereignty principles into practice (e.g., (Hellmeier *et al.*, 2023), and how appropriate assurances can be given to actors about how and to what extent data sovereignty can be operationalised by a specified data space (Heeß *et al.*, 2024).

- From an information systems perspective, there are examples of models and frameworks (e.g., Abbas et al., 2024; von Scherenberg et al., 2024) that aim to identify key aspects of data sovereignty and represent the relationships between them — with the intention of helping to guide the operationalising of data sovereignty.

As previously mentioned, the key points listed above can be used to inform our methodology for consulting with key stakeholders (section 3.2). This consultation aims to provide further insight into the meaning of data sovereignty from different stakeholder perspectives, and to identify different events in complex data lifecycles and how these events affect sovereignty in the context of data spaces (as part of T3.1).

4 THE EU REGULATORY LANDSCAPE CONCERNING DATA

Since 2018, there have been various new data-related legislation proposed and enacted in the EU (e.g., (Riis, 2023)), including the General Data Protection Regulation (GDPR), the Data Act, the Data Governance Act etc.[9] Given this expansion, Riis (2023) maintains that EU data law should be considered as an "autonomous legal field" with five key objectives, these are to "safeguard (i) a competitive market, (ii) fundamental rights, (iii) consumers, (iv) trustworthiness and (v) open data". This section will map and describe some of the key aspects of the current and future European data regulation landscape and their implications for the data lifecycle management and data sharing environment (T3.2). This section expands on an initial summary of the regulatory landscape and contractual framework which is contained in the DS2 D2.1[10].

[9] As a further overview, also see the infographic published by the International Association of Privacy Professionals (IAPP) —although noting this was published in 2023 — showing "the various European Union data initiatives and draft legislation" (Tielemans, 2023)

[10] D2.1 report - section 2.6, pp. 20-23; https://www.dataspace2.eu/results/deliverables

4.1 EU data regulation in dynamic evolution

From the regulatory perspective, the concurrent European data regulation landscape is in a dynamic development stage with several important horizontal data regulations in different stages of implementation. The work commenced with the European Strategy for Data (European Commission, 2020), which lays down the core aim for Europe to become a leader in data-driven society by creating a single market for data. Based on the European Strategy for Data, two core horizontal regulatory actions were taken, namely drafting of: the Data Governance Act (DGA) — which aims to create a structural framework for the data economy actors; and, the Data Act (DA) — which aims to increase access to data in a more specific manner — e.g., by regulating data generated through the use of connected products.

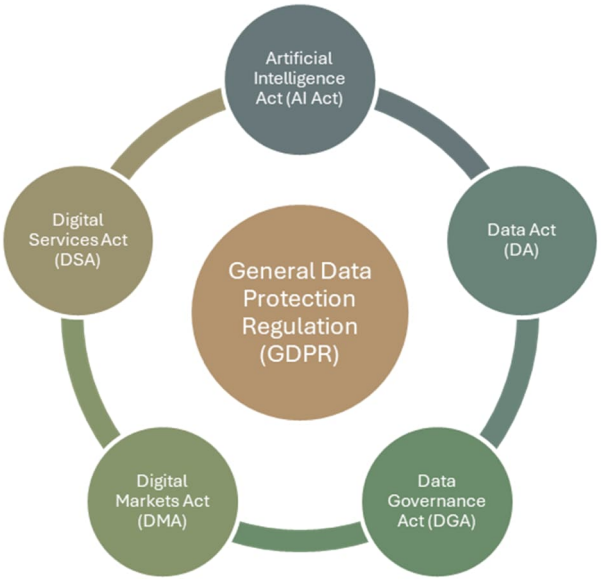

Figure 2 The Big Five Regulations of the Future European Data Economy Plus the GDPR

In addition to the DGA and DA, the European data strategy also extends to areas of consumer protection and competition law with regulations like the Digital Markets Act (DMA) and the Digital Services Act (DSA). Both regulations aim to address the imbalances in the European data economy relating to existing large digital platforms and to create a level playing field for data economy actors. The DMA regulates the gatekeepers, and the DSA regulates very large online platforms and search engines.

Together with the Artificial Intelligence Act (the AI Act), the DGA, DA, DMA and DSA form the so-called "Five Big" regulations of the future European data economy (Bräutigam *et al.*, 2022). However, there is an additional central regulation that affects the implementation of the Big Five regulations, namely the General Data Protection Regulation (GDPR). The GDPR, covering processing of personal data, has been applied already for quite some time and is deeply interlinked with all Big Five regulations.

In addition to the text of the regulations and their recitals, this study takes also into account the guidance given by the regulatory authorities (European Commission, 2024a) and (European Commission, 2024b), and studies about data intermediaries (European Commission. Joint Research Centre., 2023) and (Bobev et al., 2023).

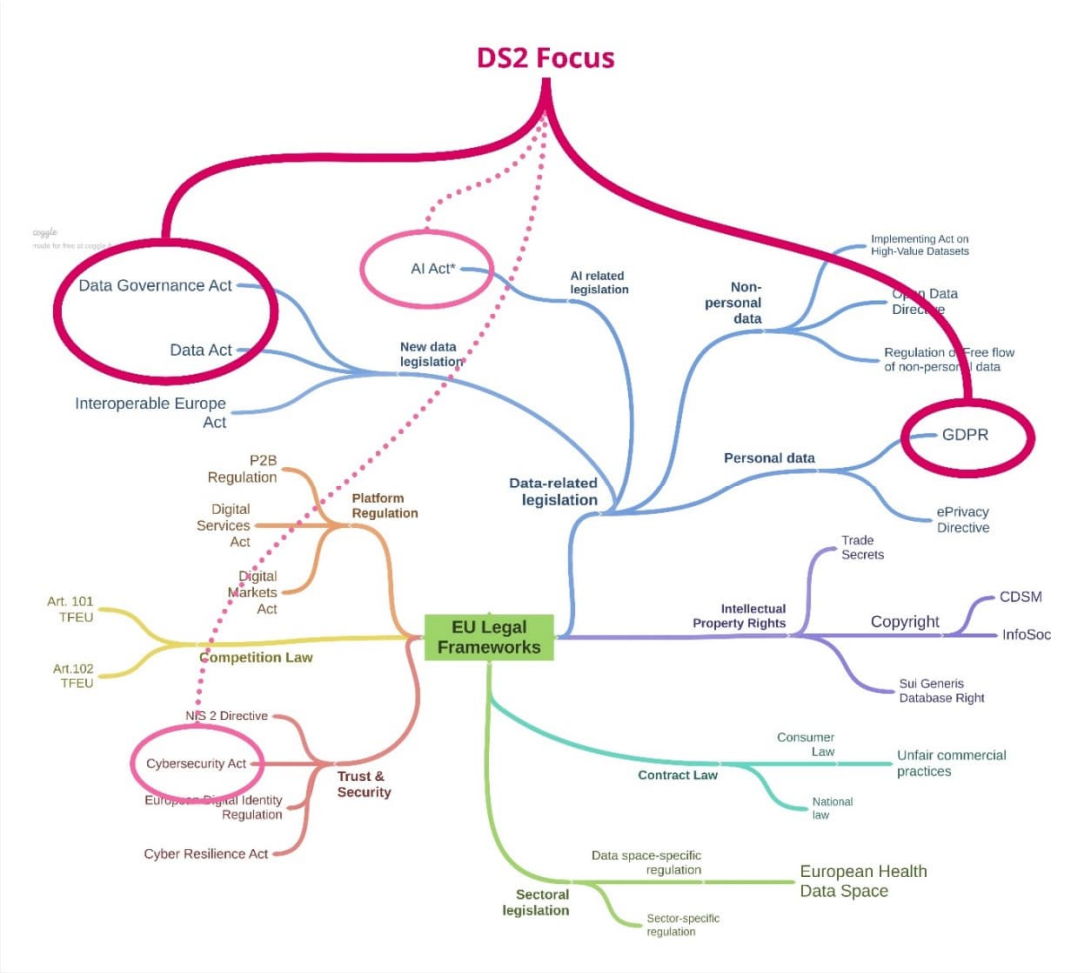

Figure 3 Key EU Data Frameworks and Areas of Focus for the DS2 Project

4.2 Mapping regulatory requirements to the DS2 project

Based on its preliminary analysis, DSSC has made a sector-agnostic mapping of the legislation that could apply to data spaces. In the context of DS2 project, the initial scoping to the DSSC mapping was undertaken during the identification of the problem space[11]. The GDPR, the DGA, the DA and the AI Act were chosen as the focus in the DS2 project given their horizontal and concurrent nature and features present in the DS2 modules and services.

It should also be acknowledged that EU data law extends further than the Big Five Regulations plus the GDPR to other legal instruments such as, the ePrivacy Directive, the Regulation on Non-Personal Data, the Open Data Directive and the Cybersecurity Act etc. (Bräutigam et al., 2022; Tielemans, 2023). To illustrate some of these key EU data frameworks, the following diagram is provided — i.e., covering the fields of data-related legislation, competition law, contract law, intellectual property law, platform regulation, sectoral legislation, and trust and security — the principal areas of focus for the DS2 project are also highlighted in Figure 3.

[11] see DS2 D2.1 report: https://www.dataspace2.eu/results/deliverables

4.3 Aims of the core regulations

The following sections will cover a general outline and key aspects from the chosen regulations within the scope of DS2. This lays down the horizontal regulatory landscape that is currently transforming the European data economy. The landscape depicted in these regulations is still in its formation and expresses the future operating environment of the existing big data platforms, data spaces and other data sharing business.

4.3.1 Data Governance Act (DGA)

DGA has entered into force 23rd June 2022 and is applicable as of 24th September 2023. Regarding its Chapter III (Data Intermediation Services), the application starts 24th September 2025.

Underlying key takeaway from the DGA is that data cannot be owned. Instead, several, differing and even contradictory rights relating to data may exist. In this kind of situation, data management and governance become an important mechanism for organizations to handle the risks and opportunities relating to data and data sharing. One core aim of the DGA is to increase trust in voluntary data sharing.

The core scope of the DGA covers:

- Reuse of certain public sector data: Mechanisms to facilitate the reuse of certain public sector data that cannot be made available as open data, e.g. health data.

- Data Intermediaries: Alternative to current big tech platforms, e.g. neutral data marketplaces, data trusts, product information management, data co-operatives

- Data altruism: Mechanisms for independent, not-for-profit organizations seeking to support objectives of general interest by making available relevant data based on data altruism at scale.

Re-use of certain categories of data held by public sector bodies continues from the basis laid down by the Open Data Directive[12] (ODD). The regulation does not force making data open, but in case opening data, processes in Chapter II apply. For the purposes of reuse, it introduces new data categories, e.g., commercially confidential data, statistically confidential data, IPR protected data and personal data.

Data Intermediation services are presented as the European alternative to current big tech platforms. The core idea behind the regulation is to provide neutrality in the intermediation of data at scale and open markets for service providers building their business on shared data. Data intermediation service providers (DISPs) are under prior notification requirement and are registered in an EU wide register. Their obligations cover, e.g., neutrality, data cross-usage prohibition, structural unbundling, authorization of add-on services, FRAND-terms, business continuity, interoperability, security, obligations to international transfers. The core definition relating to Data intermediation services can be found in DGA art 2(11):

> *Data intermediation service means a service which aims to establish commercial relationships for the purposes of data sharing between an undetermined number of data subjects and data holders on the one hand and data users on the other, through technical, legal or other means, including for the purpose of exercising the rights of data subjects in relation to personal data, excluding at least the following:*

[12] (*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)*, 2019)

a) services that obtain data from data holders and aggregate, enrich or transform the data for the purpose of adding substantial value to it and license the use of the resulting data to data users, without establishing a commercial relationship between data holders and data users;

b) services that focus on the intermediation of copyright-protected content;

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

d) data sharing services offered by public sector bodies that do not aim to establish commercial relationships

Regulation for Data altruism is based on voluntary registration on national registries. For such Data altruism organizations, the regulation introduces obligations relating to, e.g., transparency, annual activity reporting, safeguards, data cross-usage prohibition, consent/permission tools, and security.

Additional parts of the DGA cover the activities of the European Data Innovation Board (EDIB) in implementing the DGA and advising and assisting the EU Commission, and terms relating to international access and transfers.

Further information can be found, e.g., from these sources:

- Open data | Shaping Europe's digital future (*From the Public Sector Information (PSI) Directive to the Open Data Directive | Shaping Europe's digital future*, 2023)

- European Data Governance Act | Shaping Europe's digital future *(European Data Governance Act | Shaping Europe's digital future, 2024)*

- Data Governance Act explained | Shaping Europe's digital future *(Data Governance Act explained | Shaping Europe's digital future, 2022)*

4.3.2 Data Act (DA)

The European Data Act (DA) is a new regulation that aims to facilitate data sharing across different sectors by removing existing barriers to data access and use. The DA has entered into force 11th January 2024 and the application of most of its clauses starts 12th September 2025. The DA aims to boost innovation by removing barriers obstructing consumers and businesses from the access to data. It will allow a wider range of private and public entities to share data. The legislation will clarify who can access data, on what basis and under what conditions.

DA covers a wide range of activities in relation to business-to-consumer (B2C), business-to-business (B2B), and business-to-government (B2G).

In the DA, different areas of regulation are covered as follows:

- B2C and B2B Connected products Ch. II

- B2B Obligations pursuant to Union law Ch. III
- B2B Unfair contractual terms Ch. IV
- B2G Exceptional need Ch. V
- Switching between data processing services Ch. VI
- International governmental access and transfer of non-personal data Ch. VII
- Interoperability Ch. VIII
- Implementation Ch. IX

One of the core areas of regulation covers the usage data of connected products (e.g., from IoT device). DA is based on principles of access-by-design and interoperability. Access-by-design covers data generated by a connected product or generated during the provision of related services. This access is granted to data subjects, users or to data recipients designated by those users on the request of the user or data subject.

The access ensures that the users can use the connected product data, including by sharing them with third parties of their choice. This voluntary data sharing to third party service providers is based on compensation on FRAND terms. Compensation on FRAND terms, ensure making data available to data recipients in the Union on fair, reasonable and non-discriminatory terms and in a transparent manner. Reasonable compensation is also allowed in cases the data holders are legally obliged to make data available to the data recipient in business-to-business relations. In such cases, a data holder, a data recipient or a third party should not directly or indirectly charge consumers or data subjects a fee, compensation or costs for sharing data or accessing it.
Additionally, DA facilitates the switching between data processing services, provides fair contractual terms for SMEs, covers business-to-government data sharing, international safeguards and protection of personal data.
For SMEs, the DA aims also to provide fair contractual terms. Start-ups, SMEs and companies from traditional sectors with less-developed digital capabilities struggle to obtain access to relevant data. DA aims to facilitate access to data for these entities, while ensuring that the corresponding obligations are scoped as proportionately as possible to avoid overreach. When organizations draw up their data sharing contracts, the law will rebalance the bargaining power in favour of SMEs, protecting them from unfair contract terms imposed by organizations in a much stronger bargaining position.

To balance the interests of the stakeholders, DA contains strengthened provisions to protect trade secrets and avoid a situation where increased access to data is used by competitors to retro-engineer services or devices. DA also lays down the development of interoperability standards and common specifications for such data.
Under the rules of business-to-government data sharing, the DA covers making available of data to public sector bodies or Union institutions, agencies or bodies, where there is an exceptional need in the public interest.

DA introduces safeguards against unlawful international governmental access to non-personal data and states that Union law on the protection of personal data, privacy and confidentiality of communications and integrity of terminal equipment shall apply to any personal data. Any contractual term in a data sharing agreement between data holders and data recipients which, to the detriment of the data subjects, undermines the application of their rights to privacy and data protection, derogates from it, or varies its effect, will be void.
Further information can be found, e.g., from these sources:

- Data Act | Shaping Europe's digital future (*Data Act | Shaping Europe's digital future*, no date)

- Data Act explained | Shaping Europe's digital future (*Data Act explained | Shaping Europe's digital future*, no date)

While DA obligations apply more broadly than EU data spaces, "the EU data strategy envisions strong synergies between the Data Act and EU Data Spaces, which are expected to mutually reinforce each other" (Deloitte, 2024). One key area of focus for the DS2 project are the interoperability requirements set out in Chapter VIII of the DA, which are aimed at "participants in data spaces that offer data or data services to other participants" (Article 33 of the DA). For instance, the "essential requirements regarding interoperability of data, of data sharing mechanisms and services, as well as of common European data spaces" are contained in Article 33(1) and necessitate description of the following:

- "The dataset content, use restrictions, licences, data collection methodology, data quality and uncertainty [...]" (see Article 33(1)(a));

- "The data structures, data formats, vocabularies, classification schemes, taxonomies and code lists, where available [...]" (see Article 33(1)(b));

- "The technical means to access the data [...]" and "[...] their terms of use and quality of service [...]" (see Article 33(1)(c)); and,

- "Where applicable, the means to enable the interoperability of tools for automating the execution of data sharing agreements, such as smart contracts shall be provided" (see Article 33(1)(d)).

It should also be noted that Recitals 90 and 100 of the DA refer to the international standard ISO/IEC 19941:2017 Information technology — Cloud computing — Interoperability and portability (ISO, 2023): "The ISO/IEC 19941:2017 is an important international standard constituting a reference for the achievement of the objectives of this Regulation, as it contains technical considerations clarifying the complexity of such a process" (Recital 90 of the DA). However, Schneider et al. (2024) highlight that interoperability needs to go beyond technical considerations, as is illustrated by the European Interoperability Framework (European Commission: Directorate-General for Digital Services, 2017) that centres on "legal", "organisational", "semantic" and "technical" interoperability layers.

4.3.3 General Data Protection Regulation (GDPR)

The General Data Protection Regulation (GDPR) (http://data.europa.eu/eli/reg/2016/679/oj) has been applicable since 25 May 2018 (EDPS, n.d.).

The GDPR sets out obligations for controllers and processors who are processing personal data and rights to be exercised by data subjects. Seven data protection principles should be at the core of approaches to personal data processing (Information Commissioner's Office [ICO], n.d.) — these are "Lawfulness, fairness and transparency", "Purpose limitation", "Data minimisation", "Accuracy", "Storage limitation", "Integrity and confidentiality" and "Accountability" (see Article 5 of the GDPR).

1. Lawfulness, Fairness, and Transparency: Businesses must ensure that any data shared is done so lawfully and that data subjects are informed about how their data will be used. This includes providing clear information about the purpose of data processing and obtaining consent where necessary.

2. Purpose Limitation: Data collected for one purpose should not be used for another incompatible purpose. In B2B contexts, this means that companies must clarify the specific reasons for sharing personal data and ensure compliance with those stated purposes.

3. Data Minimization: Organizations should only share data that is necessary for the intended purpose. This principle encourages businesses to evaluate what personal data is essential for their operations and limit sharing accordingly.

4. Accuracy: Businesses are responsible for ensuring that the personal data they share is accurate and kept up to date. This is particularly important in B2B relationships where decisions may be based on shared data.

5. Storage Limitation: Personal data should not be retained longer than necessary. Companies must implement policies to regularly review and delete unnecessary or outdated personal data from their systems.

6. Integrity and Confidentiality: Organizations must take appropriate security measures to protect shared personal data from unauthorized access or breaches, which is crucial in maintaining trust between business partners.

7. Accountability: Businesses are required to demonstrate compliance with GDPR principles, which includes maintaining records of processing activities and being prepared to show how they uphold these obligations.

From a data protection perspective, the need for privacy-by-design and risk management and impact assessment of risks to fundamental rights and freedoms of individuals specifically related to data processing operations in data spaces has been highlighted by e.g., Agencia Española de Protección de Datos (AEPD, 2023) — the Spanish Data Protection Agency, and the European Union Agency for Cybersecurity (ENISA, 2024). Further, in terms of guidance issued by the European Data Protection Board (EDPB) and European Data Protection Supervisor (EDPS), two EDPB-EDPS Joint Opinions have been adopted on: the Proposal for a Regulation on the European Health Data Space (Joint Opinion 03/2022); and the Data Governance Act (DGA) (Joint Opinion 03/2021).

Further readings on the relevance of GDPR in B2B data sharing context can be: https://gowlingwlg.com/en/insights-resources/articles/2023/data-unlocked-data-protection-and-cyber-security and https://usercentrics.com/knowledge-hub/how-does-gdpr-affect-b2b-sales/

4.3.4 Artificial Intelligence Act (AI Act)

The Artificial Intelligence Act (AI Act) (http://data.europa.eu/eli/reg/2024/1689/oj) entered into force on 1 August 2024 (European Commission, 2024). The major start of the application is from 2nd August 2026 with the following exceptions: prohibited AI practices (2nd February 2025), general purpose AI rules (2nd August 2025), and high-risk AI systems (2nd August 2027).

Organisations will need to determine whether any of their processing activities as part of the specified data lifecycle and data sharing environment will involve the use of any AI systems as defined in the AI Act. The term

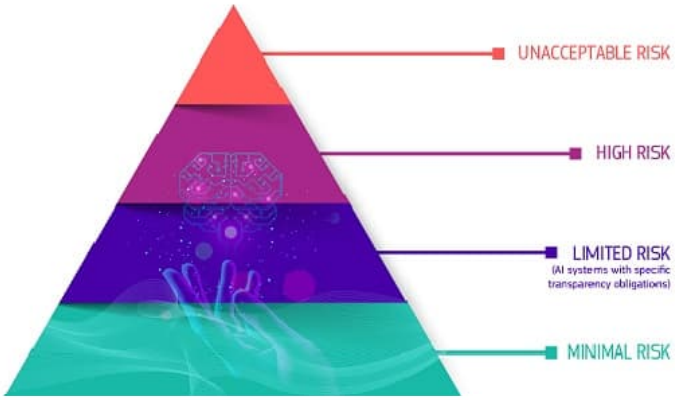

Figure 4 The risk-based approach of the AI Act

AI system is defined by Article 3(1) as *"a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments"*. However, this may be challenging as "defining AI Systems can be beguilingly complicated" (Hickman & Ja, 2024) in terms of legal definition of AI system provided.

The AI Act takes a risk-based approach setting out four levels of risk for AI systems, which are "unacceptable risk", "high risk", "limited risk" and "minimal risk". The term risk is defined by the Article 3(2) as "the combination of the probability of an occurrence of harm and the severity of that harm".

It should also be noted that data spaces are referred to in Recital 68 of the AI Act: "[…] European common data spaces established by the Commission and the facilitation of data sharing between businesses and with government in the public interest will be instrumental to provide trustful, accountable and non-discriminatory access to high-quality data for the training, validation and testing of AI systems […]"

The AI Act has entered into force 1st August 2024. The major start of the application is from 2nd August 2026 with the following exceptions: prohibited AI practices (2nd February 2025), general purpose AI rules (2nd August 2025), and high-risk AI systems (2nd August 2027).

The AI Act is a risk-based regulation and resembles the regulation relating to product safety, e.g. CE-markings and notified bodies. Basic actors are providers and deployers and main responsibilities are vested for them. The control can either be internal by the provider or external by a conformity assessment body. The core AI Act Definitions relating to this can be *found* in AI Act art 3 (3), (4) and (20):

> *''provider' means a natural or legal person, public* authority*, agency or other body that develops an AI system or a general-purpose AI model or that has an AI system, or a general-purpose AI model developed and places it on the market or puts the AI system into service under its own name or trademark, whether for payment or free of charge;*
>
> *'deployer' means a natural or legal person, public authority, agency or other body using an AI system under its authority except where the AI system is used during a personal non-professional activity;*
>
> *'conformity assessment' means the process of demonstrating whether the requirements set out in Chapter II, Section 2 relating to a high-risk AI system have been fulfilled;*

Taking a closer look at the technical concept underlying the AI Act, it is possible to compare the AI system definition of the AI Act art 3(1) and the OECD and observe them being quite close to each other.

AI Act Definition:

> *'AI system' means a machine-based system designed to operate with varying levels of autonomy, that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.*

OECD definition (*AI Principles Overview*, no date):

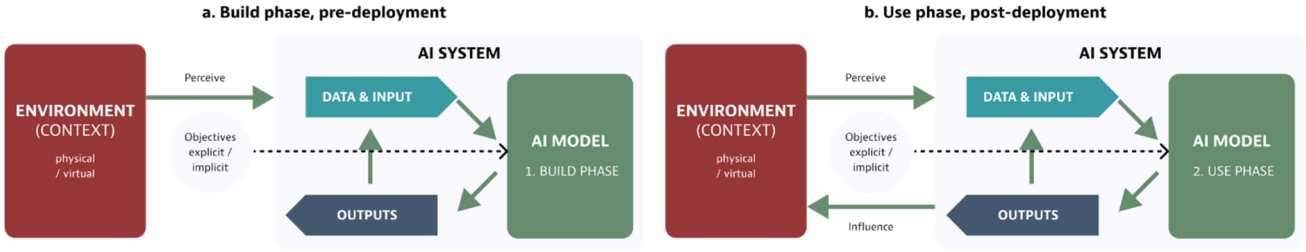

Figure 5 OECD definition of an AI system ("AI Principles Overview,")

Differences in the interpretations emerge when looking in more detail into further definitions. These definitions start to show also the political goals behind the regulation. This ambiguity can be monitored for instance in the AI Act Definition of General-purpose AI model (which is close to the previously used term of a Foundation model) in AI Act art. 3(63).

> *'general-purpose AI model' means an AI model, including where such an AI model is trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications, except AI models that are used for research, development or prototyping activities before they are released on the market;*

The definition contains also a distinction between conventional GPAI models and GPAI models with systemic risks. The ambiguity becomes even greater when we talk about AGI (Artificial General Intelligence).

In no way is the European Union alone in regulating the AI. Similar efforts to lay down some boundaries for the development and deployment of AI has emerged all over the world, e.g., US, China, Canada, South-Korea, Brazil. Also, several frameworks are active (e.g., NIST, OECD, UNESCO, G7 Hiroshima Process) and technical standards are drafted (e.g. ISO, IEEE, CEN-CENELEC).

Further information can be found, e.g., from these sources:

- AI Act | Shaping Europe's digital future*(AI Act | Shaping Europe's digital future, 2024)*
- AI Principles Overview - OECD.AI (*AI Principles Overview*, no date)

4.4 Reflecting the regulatory landscape from the perspective of DS2

Deliverable D2.1 has identified certain initial regulatory requirements and enablers from the perspective of DS2:

- GDPR:
    - DS2 can support the fulfilment of GDPR requirements by providing modules related identity management and traceability of data

- DGA:
    - DS2 related services may fall under data intermediary services, and thus, it has implications for the data space governance authority and participating organisations (increasing trust in the data sharing, making more data available on the market, neutral data intermediaries)
- DA:
    - DS2 may provide tools and modules for data interoperability, data models and data exchange.
    - DS2 modules need to maintain DA integrity
    - DS 2 modules can help to achieve the goals of the DA (access control and interoperability)
- AI Act:
    - DS2 project can provide functionalities that helps to achieve the AI act goals (trustworthiness, respect for fundamental rights, safety and ethical principles)

Use case descriptions (see deliverable D2.2) have identified using IoT-sensors and AI-tools. They have also described the need for creating multi-stakeholder data sharing collaboration and ecosystems. Need for facilitation and use of analytics has also been identified. These are all affected by the regulations identified above:

- Access to IoT-sensor data is one of the core areas of the DA, with the aim of increasing access to data enabling the flow of data from a data provider to a data consumer/user.
- AI-tools to be developed are covered by the AI Act.
- Multi-stakeholder environments and ecosystems require intermediators, operators and orchestrators and those actors and roles are subject to the requirements of the DGA.
- Data sharing environments also require service providers either for providing access to data (upstream service providers) or providing services built on top of data (downstream service providers). Both DGA and DA affect the environment such service providers operate in.

5 DATA LIFECYCLE MANAGEMENT

Data Lifecycle Management is a strategic approach that tracks and manages data's transformative journey from raw information to valuable data products, encompassing stages of generation, enhancement, processing, sharing, and potential reuse. As noted by Ball (Ball, 2012), lifecycle models provide a structured framework for understanding how data can be systematically developed, refined, and leveraged to create actionable assets across different contexts and stakeholder needs.

5.1 Building opportunity-based data lifecycle management

DS2 has identified the possibilities embedded in creating an opportunity-based data lifecycle management framework for data space participants. This serves the needs of the DS2 use cases in terms of building the

implementations on business opportunities and drafting of a reflecting smart contractual structure. The framework covers the processes and tools (e.g. concepts, guidelines, canvases, visual mapping tools, templates, examples) for any data space participants to build their own data product related data lifecycle management with the core aim of addressing cross-data-space opportunities. The framework also addresses interlinkages that these processes and tools have on the technological layers, for instance to the enforcement of policies. Covering all these layers, partly on the side of the human in the loop and partly on the implementation of technology, is needed to make decisions whether to implement technological solutions for smart contracting or supporting structures for human-in-the-loop.

5.1.1 Need for supporting structures

As shown below, the big idea of EU prosperity from data is a topic of legislative debates involving culture-bound social choices, which have so far led to several new laws that have been outlined in the preceding section. These laws can inform the formulation of new regulations. Also, the new laws can be related to the many existing regulations that are relevant to the management of data dynamics, such as ISO/IEC 27001:2022 Information security, cybersecurity and privacy protection — Information security management systems — Requirements.

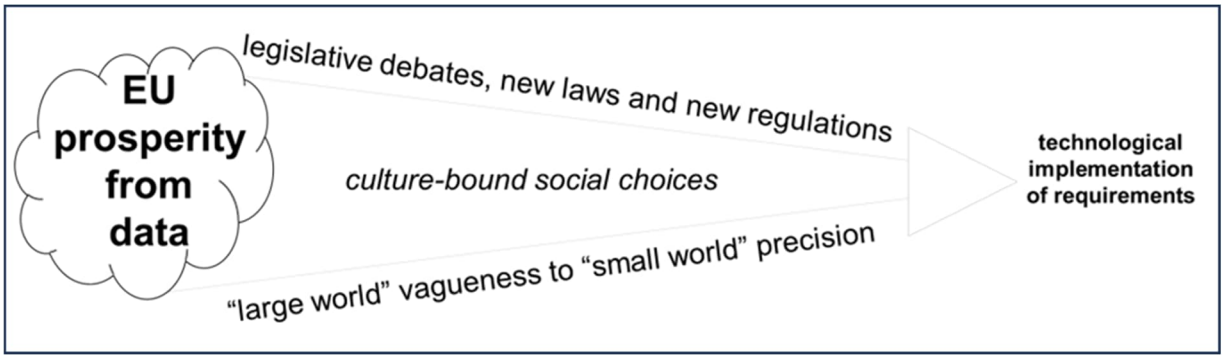

Figure 6 From the big idea to technological implementation of requirements [adapted from (Fox and Rey, 2024)]

The progression of culture-bound social choices from a big idea to technological implementations of related requirements involves progression from the intractable unpredictability of vaguely defined "large worlds" to more tractable, more predictable, more precisely defined "small worlds" (Binmore, 2017) in which specific regulations can be implemented. However, different people can have different beliefs that entail different world views. Such differences are summarized with scientific terms such as motivated cognition (Jost *et al.*, 2018), motivated reasoning (Druckman and McGrath, 2019), and motivated numeracy (Nurse and Grant, 2020), which can be summarized colloquially with phrases such as "we don't see things as they are, we see them as we are" (quoteinvestigator.com/2014/03/09/as-we-are/; Dunning and Balcetis, 2013). Thus, rather than there being only one possible encoding for one regulation and one possible encoding for one requirement, there can be different encodings by different people of the same regulation and of the same requirement. This can happen even when carefully prepared everyday language is used (Gros, Thibaut and Sander, 2021).

Accordingly, while regulations and requirements provide bases for each specific dataspace implementation, it should not be assumed that all those involved in a specific dataspace implementation will see regulations and requirements in the same way. Rather, as with any other information, even information that is professionally prepared and presented (Perloff, 2015), different people can see regulations and requirements in different and even opposing ways. As DS2 is focused on complex lifecycles of cross-sector data exchanges, it can be expected that there will often be different and even opposing views of regulations and requirements. Thus,

no matter how diligent the formulation of regulations and requirements, there can be loopholes in their implementation. Loopholes involve misalignments between what should be done and what is actually done in practice (Katz, 2010), (Licato and Marji, 2018).

Loopholes can arise from different people seeing regulations and requirements differently to the people who formulated the regulations and requirements, because they have different beliefs to the people who formulated them. In other words, the same semantic encodings can have different personal encodings in pragmatics (Bach, 1997; Goodman and Frank, 2016). Accordingly, to reduce potential for loopholes, semantic encodings and encodings in pragmatics need to be aligned during the process of formalizing the consensus between parties. This can be done through collaborative application of standard techniques such as adversarial safety analysis (Johnston, 2004) and scenario-based failure mode and effects analysis (Kmenta and Ishii, 2004) as well as the SPYDERISK tool being deployed in this project or the opportunity-based approaches.

5.1.2 Opportunities from data spaces

However, as well as closing potentially undesirable loopholes, supporting structures can encompass the ideation of new opportunities among different people who have different world views based on different beliefs. As summarized in ., opportunities from new technologies can be considered in terms of automation opportunities, informational opportunities, transformational opportunities, and interactions between opportunities (Mooney, Gurbaxani and Kraemer, 1996; Fox, 2016). Automation opportunities and informational opportunities include new types of productivity based on the substitution of human labour and the availability of more information for decision making. By contrast, transformational opportunities can include new markets being created.

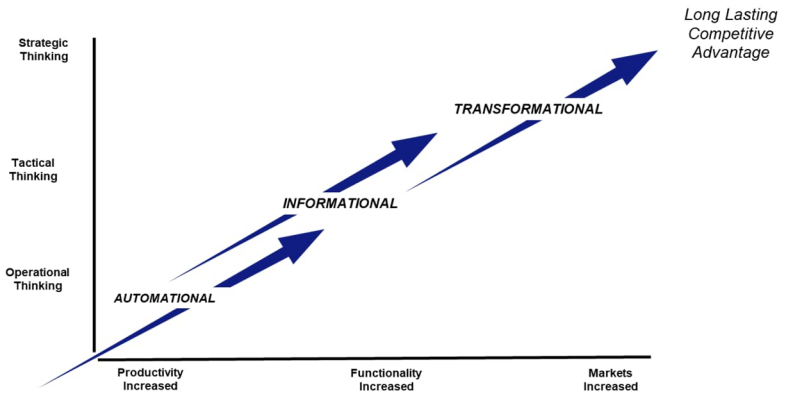

Figure 7 Relationships between automation, informational, and transformational opportunities

Automation opportunities and informational opportunities from new technologies tend to be widely reported when new technologies are being hyped. However, the reproducibility of automation opportunities and informational opportunities limits their potential to differentiate organizations in competitive markets. By contrast, the achievement of transformational opportunities depends upon ideating entirely new ways of doing things that cannot easily be copied. A major challenge in ideating entirely new ways of doing things that cannot be copied is to dismantle preconceptions that are based on established organizational practices: for example, organizational practices that predate data spaces. Methods for dismantling preconceptions can be within the scope of supporting structures (Fox, 2019).

5.1.3 DS2 opportunity-based method outline

The method to be used to address the challenges embedded in loopholes and opportunities with the aim for creating appropriate supporting structures is divided into 5 themes:

- Theme 1: Data lifecycles in DS2
- Theme 2: Regulatory enablers relating to DS2
- Theme 3: Business cases and business models of the DS2 participants
- Theme 4: Data products in DS2
- Theme 5: Implementing enabling technology and DS2 modules

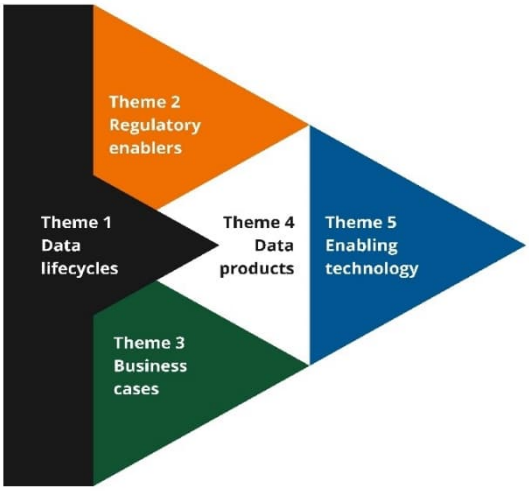

Figure 8 Five themes of the DS2 Opportunity-based approach

Each of the themes investigates one core aspect relevant for data sharing in DS2. Theme 1 aims to set out the basic context, i.e., data lifecycles in data spaces and DS2. Based on this, further steps can be scoped. Additionally, it helps in laying down what should be implemented through automated data management (technical implementations, smart contracting, etc.) and what should be left for the supporting structures (guidance, advice, human-readable contracting, etc.). Theme 1 will be based on the following layered structure:

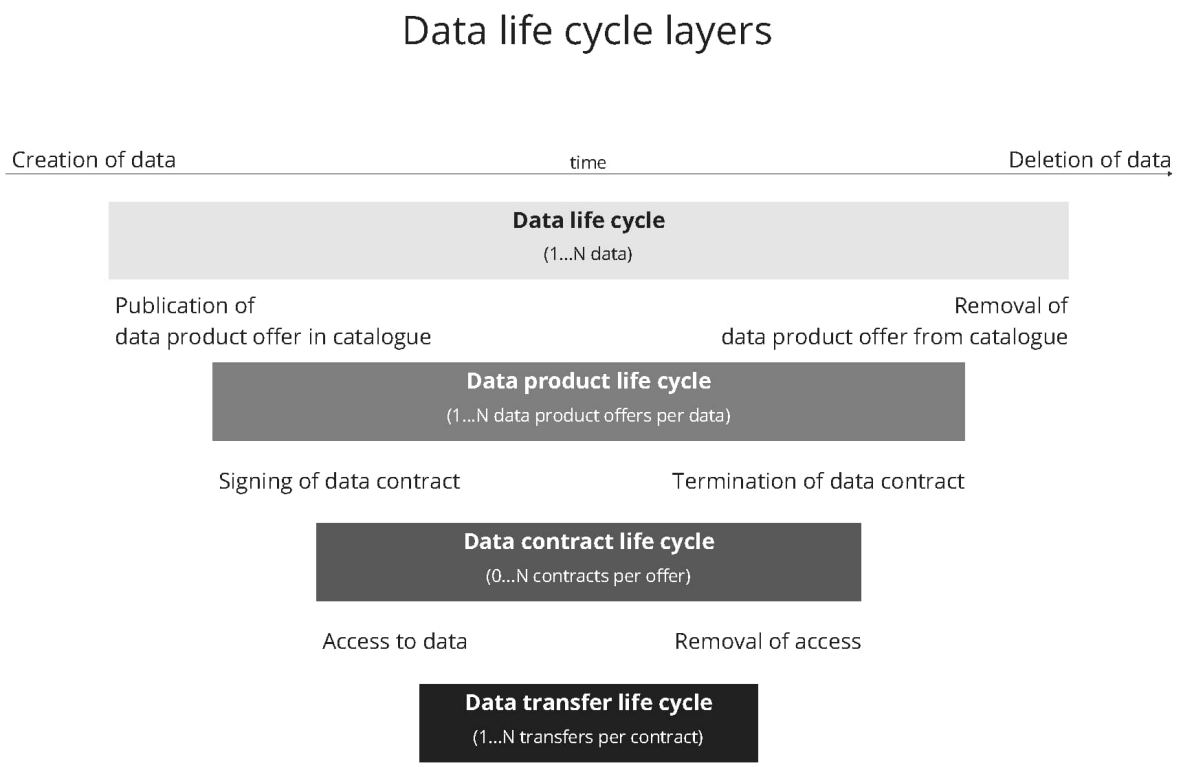

Figure 9 Data life cycle layers

The opportunity-based method has its roots in the DSSC Building blocks (DSSC, 2024). The most relevant DSSC Building blocks affecting the data life cycles have been identified in the following figure. The opportunity-based method continues from the above mentioned DSSC Building blocks with its Themes 2-4. Theme 2 investigates the evolving surrounding regulatory environment to find requirements and enablers for DS2 data sharing aiming to seize future opportunities. Theme 3 delves into the DS2 use cases from the perspective of individual DS2 participants to find business opportunities. Theme 4 aims to transform the business opportunities into DS2 data products, which can then be further implemented into DS2 technological components (e.g., machine-readable policies). In the methodology, Theme 4, Data products in DS2 plays an essential role as the interlinking building block between the borderline of automation opportunities and other types of opportunities (informational and transformational). Figure 4.1.2.3.2 expresses these themes, the opportunity-based approach and relevant DSSC Building blocks in relation to the data life cycle layers.

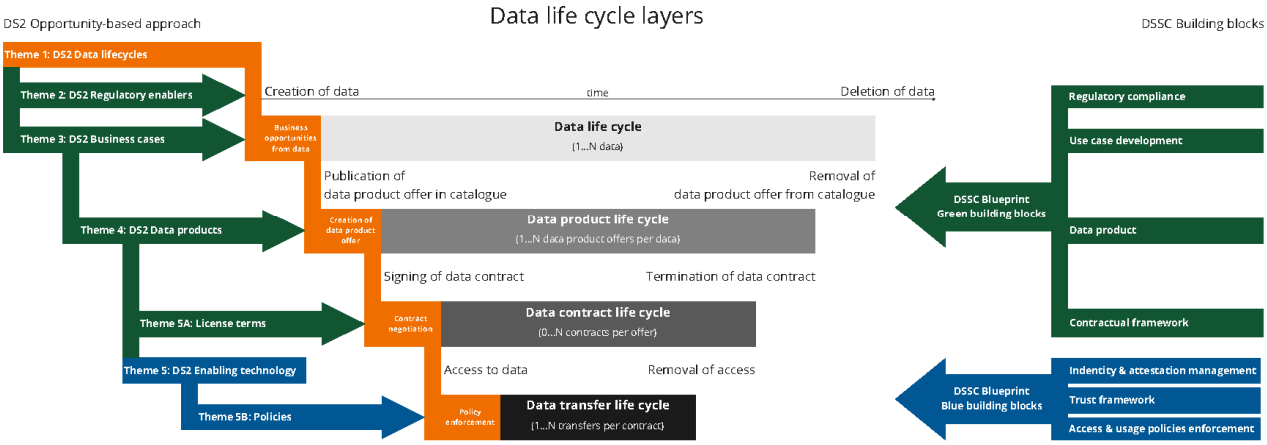

Figure 10 Data life cycle layers, the DS2 opportunity-based method and relevant DSSC Building blocks

5.2 Expected outcome

By using this methodology, it is possible to start identifying the DS2 modules and services affected and then the risks imposed on DS2 modules and services but also, at the same time the new opened opportunities. Combining Risk and opportunities this methodology supports SWAT analysis of the adoption of DS2 modules in B2B data sharing scenarios.

6 DEFINING AND BUILDING TRUST IN DATA SPACES

Enabling "secure and trustworthy data transactions between participants while supporting trust" is another core aspect of data spaces, as set out in the data space definition provided by DSSC) followed by the DS2 project. Further, the notion of trust appears as a prominent feature of a European Strategy for Data[13] with the majority of data-related legislation highlighting the significance of trust (Riis, 2023).[14] It is worthwhile to note that the question of "[to] what extent can regulation influence trust?" is one recently explored by (Tamò-Larrieux *et al.*, 2024) in relation to AI. More specifically, in terms of B2B data sharing and re-usage trust is highlighted as a key aspect (e.g., (Richter and Slowinski, 2019)). For instance, according to a European

[13] Take for instance, the key goals of the European strategies for data and AI where the intention is for Europe to be considered as "a trusted digital leader" — and more specifically a "leader in trustworthy Artificial Intelligence" and the "data economy". "Common European Data Spaces will make more data available for access and reuse. This will be done in a trustworthy and secure environment for the benefit of European businesses and citizens." (Quote from European Commission website: https://digital-strategy.ec.europa.eu/en/policies/data-spaces) "Trust is a fundamental enabler for the success of the European Health Data Space. EHDS will provide a trustworthy setting for secure access to and processing a wide range of health data." Quote from European Commission website : https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en)

[14] How is trust referred to as part of legislation? E.g., Recital 102 of the Data Act (https://eur-lex.europa.eu/eli/reg/2023/2854) states: *"To foster further trust in data, it is important that safeguards to ensure control of their data by Union citizens, the public sector bodies and businesses are implemented to the extent possible. In addition, Union law, values and standards regarding, inter alia, security, data protection and privacy, and consumer protection should be upheld. [...]."* Recital 7 of the GDPR states: "Those developments require a strong and more coherent data protection framework in the Union, backed by strong enforcement, given the importance of creating the trust that will allow the digital economy to develop across the internal market. [...]." (https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32016R0679)

Commission et al. (2018) report concerning a study on data sharing between organizations in Europe, "building trust with data users and suppliers" is presented as one key aspect of supporting "companies to successfully share data in business contexts". Against this backdrop, meanings and interpretations of trust in the data space context are explored (section 6.1). The four approaches used in DS2 are then described where the intention is to demonstrate trustworthiness in data spaces — i.e., risk assessment (section 6.2), opportunity-based data lifecycle management (section 6.3), policy and contract management enforcement (section 6.4), and logging processes (section 6.5).

6.1 A need to demonstrate trustworthiness

In the words of (Marsh *et al.*, 2020): "Trust matters because we use it to make decisions about things". "Well-placed trust" is considered as crucial for fostering data interactions in socio-technical systems (Thornton, Knowles and Blair, 2022) such as data spaces. As (O'Neill, 2018) states: "Trust is valuable when placed in trustworthy agents and activities, but damaging or costly when (mis)placed in untrustworthy agents and activities". Consider the important distinction made between trust and trustworthiness by (O'Hara, 2012)):

> *"Trust is an attitude that one takes to the trustworthiness of another; in turn, the other's trustworthiness is a property that they have. Broadly speaking, it is the property that they will do what they say they will do. […] [/] Trustworthiness is naturally not context-independent; one is not trustworthy in all respects in all contexts" (O'Hara, 2012).*

To build trust, data spaces therefore need to be able to demonstrate that they are "worthy of trust" (e.g., (O'Neill, 2013; Babb, 2021) — to do this, one crucial factor is that trustworthiness is evidenced and presented in meaningful ways to different actors involved in data spaces. Putting in place effective organisational and technical assurance measures is therefore a key aspect of fostering trustworthy B2B interactions within data spaces. For instance, the notion of "trustworthy and ethical assurance" is defined by (Burr *et al.*, 2024) as follows

> *"Trustworthy and Ethical Assurance is a structured approach to the communication of reasons and evidence about a data-driven technology or system, which helps stakeholders and affected users understand and evaluate the trustworthiness and validity of an argument made about some property or goal of the technology or system." (Burr et al., 2024)*

From a technical perspective, in the US National Institute of Standards and Technology NIST SP 800-160v1r1 special publication on engineering trustworthy secure systems, the notion of assurance is given as the basis for the "trustworthiness of a system" (Data Sharing Coalition, 2021) — where assurance is defined as "the grounds for justified confidence that a claim or set of claims has been or will be achieved" (Ross et al., 2021 citing ISO/IEC/IEEE 15026-1:2019.). Further, the related concept of data assurance should also be mentioned, which is defined by the Open Data Institute (ODI) as "the process, or set of processes, that increase confidence that data will meet a specific need, and that organisations collecting, accessing, using and sharing data are doing so in trustworthy ways" (Davies, 2023).

However, there is also the contention that reliance on such assurance mechanisms can result in a "trust paradox" — i.e., "because the *assurance* structures designed to make interpersonal trust possible in uncertain environments undermine the need for trust in the first place" (Cheshire, 2011); italics emphasis in original text). By way of illustration, consider the debate in the Journal of Medical Ethics over whether trusted research environments (TREs) "remove the need for trust" given the "assurances or guarantees of data privacy and security" provided (Graham *et al.*, 2023), or rather "make it easier for people to trust but there is still a need

for that trust" (Affleck *et al.*, 2023). TREs are conceived as "highly secure computing environments that provide remote access to health data for approved researchers to use in research [...]" (Health Data Research U. K. (hdr, no date).

6.1.1 Trust definitions

Designing trustworthy socio-technical systems for data access and re-use can be challenging given the elusive nature of the word trust, and the multiple meanings, views and approaches to trust across different disciplines (Thornton, Knowles and Blair, 2022). For instance, there are various types of trust including "[i]nterpersonal trust", "[i]nstitutional trust", "[o]rganisational trust", "[c]ategorical trust", "[s]ystem trust" and "[s]elf trust" (Smart *et al.*, 2021). As examples, here are some well-known definitions of trust provided in the literature. First, consider the definition of trust provided by (Mayer, Davis and Schoorman, 1995):

> *"... the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party" (Mayer et al., 1995)*

Another definition of trust from the behavioural sciences is outlined by Rousseau et al. (1998):

> *"...a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another" (Rousseau* et al.*, 1998).*

Other more technical definitions of trust fall closer to functional specifications of a system — i.e., system properties — and could be described as 'reliance' that is a dependence on the performance of a documented or regulated activity for instance compliance with legal requirements, standards, function and so on. This type of trust relies on objective, measurable criteria and is often enforced through technological solutions and regulatory frameworks. For illustration, consider one of the trust definitions contained in the US National Institute of Standards and Technology (Standards and Technology (NIST), 2018) Glossary (also see Ross et al., 2022):

> *"A characteristic of an entity that indicates its ability to perform certain functions or services correctly, fairly and impartially, along with assurance that the entity and its identifier are genuine" (NIST, n.d.).*

Further, consider the technical definition of trustworthiness of a system from a security perspective as defined by NIST Special Publication 800-37:

> *"The degree to which an information system (including the information technology components that are used to build the system) can be expected to preserve the confidentiality, integrity, and availability of the information being processed, stored, or transmitted by the system across the full range of threats and individuals' privacy" (NIST, 2018).*

Also, from a technical perspective, Akram & Ko (2014) provide a generic definition for digital trust — also recognising that this term may have distinct meanings in "different computing domains" such as, for the "semantic web", "data provenance", "secure and trust computing":

> *"trust based either on past experience or evidence that an entity has behaved and/or will behave in accordance with the self-stated behaviour" (Akram and Ko, 2014).*

Further, it should also be mentioned that from a security perspective, a distinction is often made between different types of hard, soft and hybrid trust frameworks that can be used as part of trust management for systems, such as data spaces (Akram & Ko, 2014). The term hard trust framework is used by Akram & Ko (2014) to mean "architectures that base the measurement/foundation of trust on verifiable and independently validated hardware" (Akram & Ko, 2014). Whereas the term soft trust framework refers to "trust measurement and assessment mechanisms that do not rely on trusted hardware: examples of soft trust can be reputation-, context- and content-based trust mechanisms" (Akram & Ko, 2014). Further, from a security perspective, (Jabeen and others, 2018) make a distinction between hard trust mechanisms — i.e., "cryptographical mechanisms" — centre on "technical solutions to provide secure interactions between service providers and service consumers"; and soft trust mechanisms — i.e., "non-cryptographical mechanisms" — are "context-dependent and [...] derived using individual or social control mechanisms.

6.1.2 Meanings and interpretations of trust in data spaces

The view of trust taken by data space initiatives — such as DSSC, GAIA-X and IDSA — appears to be more functional (see below) and very close to the technical definitions of trust (see above).

6.1.2.1 DSSC Data Spaces Blueprint v1.0

As highlighted in the DS2 D2.1 report, data sovereignty and trust are presented in the DSSC Data Spaces Blue Print v1.0 as one of the key technical pillars for creating a data space, which is described as "capabilities needed for the identification of participants and assets in a data space, the establishment of trust and the possibility to define and enforce policies for access & usage control" (DSSC, 2024). The data sovereignty and trust pillar compromises of three building blocks: "Identity and Attestation Management" which refers to "management of identities and attestations within a data space to ensure the reliability and integrity of participants' information"; "Trust Framework" which relates to "verification that a participant in a data space adheres to certain rules and a common set of standards"; and, "Access and Usage Policies Enforcement" which concerns "the ability to specify policies and rules within a given data space by the data space authority and the individual participants" (DSSC, 2024). Further, 58 collected standards for data sovereignty and trust are outlined by DSSC (2023), which are grouped according to these three technical building blocks listed above.

6.1.2.2 Gaia-X Trust Framework 22.10 - Release

The Gaia-X Trust Framework (Bonfiglio, 2021) foresees verifiable credentials and linked data representations as cornerstone of its future operations. Trusted information shall be retrieved in machine readable manners, and where such manners are missing, Gaia-X will define processes to translate trusted information in a machine-readable format. This is a prerequisite of federating trusted statements within the Gaia-X Ecosystem and developing mechanisms to re-assess validity of claims within the Trust Framework. The set of computable rules known as compliance process is automated and versioned. It means that this document will also be versioned.

6.1.2.3 IDSA

Trust is composed of[15]

- Roles. Each role in the International Data Spaces has certain rights and duties. For example, the Identity Provider is responsible for offering services to create, maintain, manage, monitor, and validate

[15] https://docs.internationaldataspaces.org/ids-knowledgebase/ids-ram-4/layers-of-the-reference-architecture-model/3-layers-of-the-reference-architecture-model/3_2_functionallayer)

identity information of and for participants in the International Data Spaces. More information about the roles is given in the Business Layer.

- Identity Management. Every Connector participating in the International Data Spaces must have a unique identifier and a valid certificate. In addition, each Connector must be able to verify the identity of other Connectors (with special conditions being applied here; e.g., security profiles).

- User Certification. Each participant in the International Data Spaces must undergo certification to establish trust among all participants. More information about the certification process is given in the Certification Perspective.

- Governance aspects. This is supposed to cover the non-functional specification of trust.

6.1.2.4 Different data space configurations

The design of common European data spaces "should be guided by" the following "design principles": (i) "Data control", (ii) "Governance", (iii) "Respect of EU rules and values", (iv) "Technical data infrastructure", (v) "Interconnection and interoperability", and (vi) "Openness" (European Commission, 2022). Various governance models can be used to facilitate B2B data interactions, such as "Data monetisation", "Data marketplaces", "Industrial data platforms", "Technical enablers" and "Open data policy" (European Commission et al., 2018). It should also be mentioned that other governance models are emerging not just for B2B data interactions but more widely — such as "data sharing pools", "data co-operatives", "public data trusts" and "personal data sovereignty" (Micheli *et al.*, 2020). As a further example, also see the report on legal mechanisms for data stewardship published by the Ada Lovelace Institute & AI Council (2021) discussing "data trusts, data cooperatives and corporate and contractual models".

As previously mentioned, a key aim of the DS2 project is to identify different events in complex data lifecycles and how these events affect sovereignty (as part of T3.1). In view of this, it is important to understand the *"precise distribution of function"* within a given data space — given that this specific distribution is *"directly linked to governance and risk, as it affects the dispersal of control, agency, and trust between stakeholders"* (Boniface et al., 2020). Data spaces can have various "different configurations" — six examples of which are provided by the Agencia Española de Protección de Datos (Datos (AEPD), 2023), the Spanish Data Protection Authority, and are summarised in the table below. It should also be noted that federations of data spaces can emerge (AEPD, 2023).

	AEPD (2023)16: Data Space Configuration Type	Brief description
1	"Data Space based on sharing via a central node"	In this example configuration, "all the mediation and supervision functions" are established by a "single entity".
2	"Data Space Mediator as a central hub or data marketplace"	In this example configuration, the Data Space Mediator has a more limited role focusing on specific functions—e.g., "could be limited to managing the participants, the data catalogue and the security mechanisms, among other services, while Holders and Users can access peer-to-peer data".

[16] All quotes in table from source: AEPD, 2023 (https://www.aepd.es/documento/approach-to-data-spaces-from-gdpr-perspective.pdf), pp. 35-38 (see original text for full information and diagrams)

3	'Multiple entities providing multiple functions'	In this example configuration, the data space involves various entities providing "multiple functions", including "different offerings of the same service".
4	'Facilitating direct interaction"	In this example configuration, a data space is conceived as facilitating "direct interaction between Data Subjects and Users" — e.g., "data altruism organisations".
5	"Data Space configuration without the use of data mediation (sic) services"	In this example configuration, a data space enables data users to access data held by data providers without mediators.
6	"Data Space with data access agreements between Data Subjects and Data Holders"	In this example configuration, a data space is conceived as "an infrastructure" with a mediator "to facilitate data access agreements between Data Subjects and Data Holders".

Table 2 Six example configurations of data spaces (AEPD, 2023)

6.1.3 A holistic view of trust in DS2

Complying with applicable contractual, regulatory and other legal requirements, including the use of technical enforcement mechanisms, is not enough alone for data spaces to demonstrate that they are worthy of trust or for fostering social legitimacy — in other words to "secure a social licence" (Carter, Laurie and Dixon-Woods, 2015). It should also be mentioned that a social licence can be conceived as a "policy tool" for operationalising digital self-determination (Verhulst, 2023).

When considering how to demonstrate trustworthiness, it is important to think about the properties of trustworthiness that will have most significance for the actors who are participating in or are otherwise affected by interactions within a specified data space — and for data spaces and society more generally. For instance, although specifically related to sharing patient data, consider the approach taken by Understanding Patient Data (2024) for helping to guide "organisations to demonstrate that they are trustworthy" where six main "trustworthiness characteristics" are highlighted, which are: "[m]otivation", "[c]ompetence", "[t]ransparency", "[g]overnance", and "[p]ublic [p]articipation". In a similar vein, the notion can also be considered with reference to "proxies of trustworthiness" that is, "attempts to perform trustworthiness relative to underpinning values" (Harvey and Laurie, 2024). By way of illustration, "five common proxies of trustworthiness" in human health research are provided, which are "(i) consent, (ii) anonymization, (iii) public engagement, (iv) openness, and (v) accountability" (Harvey & Laurie, 2024). As a further example, in the context of B2B relationships, (Alves, Campos and Oliveira, 2012) present a model with fifteen "determinants of the trustworthiness of a supplier" — e.g., "[p]ast and experience", "[p]roduct's quality", "[d]eadlines fulfilment", "[a]ccreditation", "[c]ooperative norms" etc.

In the DS2 project, our approach is focused on one particular aspect of demonstrating trustworthiness — that is, how and to what extent trustworthiness can be evidenced and presented in meaningful ways to different actors involved in data spaces through specified assurance mechanisms. Such mechanisms include risk assessment, opportunity-based data lifecycle management, policy and contract management enforcement, and logging processes (which are described in more detail in the following sub-sections of this report). In particular, the research is exploring how organisations can improve knowledge through collaboration and demonstrate trustworthiness to each other through risk communication— e.g., information sharing between organisations concerning system model-based risk assessment in the domain of information security.

The assumption is that each decision maker in a data space has a combination of subjective and objective requirements that needs to be met before a data access decision can be made. Assorted decision makers will have different requirements dependent on the given context and perception of the barriers to the data sharing. Some of these barriers can be lowered by mutually sharing an optimal set of information on the providers and consumers systems. One simple example from the cybersecurity domain is the cyber-essential certification that includes a set of information about the security system to be shared with potential business partners to reassure them about the quality of the security at both ends of the business relationship. By providing a risk assessment that includes key aspects relevant for the decision makers, the barriers will be lowered.

The technical components of the data space, such as the enforcement mechanisms, will work as enablers — the intention being that their presence aims to increase the trustworthiness of the distributed system by providing meaningful and effective assurance, which in turn will lower the overall risk.

6.2 Risk assessment

Risk assessment tools can be used to support organisations in making consistent and transparent governance decisions as part of data spaces, such as around data access, sharing and linkage. In some cases, such governance decisions may be referred to as "trust decisions" — i.e., "the decisions that X makes on the basis of their beliefs about Y 's trustworthiness" — which should be further distinguished from "trust beliefs" — i.e., "the beliefs that X has about Y 's trustworthiness" (all quotes: Smart et al., 2021). By way of illustration, consider the example of the Data Spaces Support Centre Blueprint v1.0 (DSCC, 2024) where risk assessment is presented as an essential component of its Trust Framework setting out the "composition of policies, rules, standards, and procedures designed for trust decisions in data spaces based on verifiable credentials". Here trust is defined as "a firm belief in the reliability, truth, or ability of a party to fulfil a promise" where "trust or distrust is the result of a threshold that is set by the decision-making party and is based on a risk assessment [...] to determine the level of trust from the evidence provided" (DSSC, 2024). Further, the meaning of trust decision is given as "a judgement made by a party to rely on some statement being true without requiring absolute proof or certainty" (DSCC, 2024).

6.2.1 Our approach using DSM

Risk assessment is therefore presented by the DSSC as one crucial aspect of demonstrating trustworthiness in relation to data spaces. In the DS2 project, our approach uses the automated risk assessment module DSM presented in D2.2. The intention being that each organisation carries out a risk assessment locally and then can share certain aspects of the risk assessment report with the aim of demonstrating trustworthiness to each other through risk communication.

6.2.1.1 Working model for demonstrating trustworthiness in data space interactions for DS2

For the purposes of illustration, the following working model provides a high-level representation of demonstrating trustworthiness through specified risk assurance mechanisms for data spaces in the DS2 project. It must be noted that the diagram provides a view of trustworthiness from an information systems perspective for the project and therefore does not include e.g., other trustworthiness properties and related ways in which data spaces and their participants can demonstrate that they are trustworthy (e.g., through collective governance mechanisms, audit etc.); or other external factors that might inform trust assessment.

6.2.1.1.1 Trustworthiness model description

The working model for demonstrating trustworthiness in data space interactions for DS2 has been developed primarily through our interpretations of the "Trust Process Model" outlined in Smart et al. (2021) together

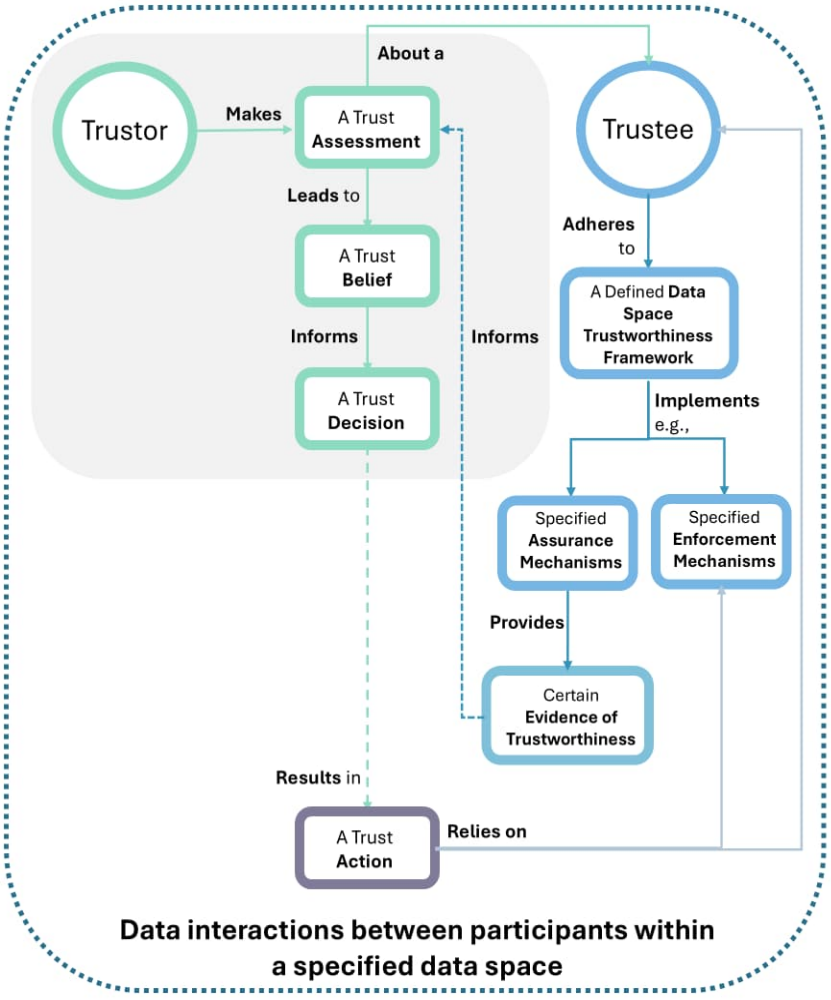

Figure 11 Working model for demonstrating trustworthiness in data space interactions for DS2 [Based on interpretations of Smart et al., 2021; DSSC, 2024; Ross et al., 2022]

with the DSSC Trust Framework (DSSC, 2024), and the approach to trustworthiness and assurance presented in NIST SP 800-160v1r1: Engineering Trustworthy Secure Systems (see Appendix F; Ross et al., 2022). These are the key elements of this model:

- An organisation which is a data provider (trustor) is required to make a decision (trust decision) about sharing some specified data with an organisation which is a data user (trustee) operating in a specified data space.

- To do this, the data provider makes an assessment about the trustworthiness of the data user (trust assessment) in the context of the data space (e.g., Smart et al., 2021).

- The data user aims to demonstrate their trustworthiness to the data provider by offering certain evidence of trustworthiness (DSSC, 2024; Ross et al., 2022). These assurance mechanisms form part of a defined data space trustworthiness framework for the specified data space. In this case, such evidence (e.g., risk reporting — such as ISO 27001 risk treatment plans) is provided from given assurance mechanisms (e.g., for risk management), which aim to help the data provider to evaluate particular trustworthiness properties (e.g., information security) in the context of the specified data space (e.g., Burr et al., 2024).

- The data provider may use the certain evidence of trustworthiness presented to inform their trust assessment. Of course, it is important to highlight that in some cases a trustor could make a trust decision without knowledge of certain evidence of trustworthiness or may not use such evidence where it is available (e.g., Ross et al., 2022) — a point which is indicated by the dashes in the arrows (see figure).

- The outcome of trust assessment is a trust belief (Smart et al., 2021).

- The data provider then makes a decision (trust decision) to share some specified data based on their beliefs (trust belief) about the data user in this context.

- A trust decision made by the data provider leads to a trust action (Smart et al., 2021) where data is shared with the data user under certain terms and conditions in the data space.

- The data provider relies on the data user to adhere to the terms and conditions of use now data has been shared with them, and the specified enforcement mechanisms implemented as part of the defined data space trustworthiness framework as further measures to ensure that these specifications are complied with.

The picture above provides an abstraction of the mechanisms already presented in D2.2. From a practical point of view, the DSM will work requiring that both the data provider and data consumer would perform a local assessment and then would share the result with the counterpart (provides "Certain evidence of trustworthiness" in the figure above). Separate modules of the DSM allow the execution of the local assessment and the comparative analysis of the results supporting the "trust assessment" in the figure above. Details of the DSM implementation can be found in D2.2.

It is important to highlight that DMS works in the preliminary phase, when the decision of sharing data needs to be taken. It supports the user in taking the decision, but it does not have an active role when the data sharing actually happens.

6.2.1.2 Future Needs

As a first practical step, the technical components of a trustworthiness framework developed or anyway present in the use cases must be identified (e.g. identity management and policy enforcements supported by blockchain based Clearing House module). Then, DS2 team will work with decision makers and users to understand what the perceived barriers to data sharing are, and how to lower them (i.e., what type of information they would need to have to decide to share certain data with other businesses). This information will be encoded in the Risk Assessment knowledge base (e.g., compliance with EU regulation). Further, we will identify the set of information (e.g., certain evidence of trustworthiness) that a business partner would be willing to share to enable the data-sharing to take place (trade-off between sensitive information and trust gain). Currently, this is envisaged to be, at least, the cyber-essential set of information. To help address any policy conflicts across dataspaces, the decision makers' requirements will be expressed in policies where this is possible. Finally, the risk assessment for the three use cases is performed.

6.3 DS2 implementation of the opportunity-based data lifecycle management

The operational implementation of DS2 is grounded in five interconnected themes which provide a comprehensive framework for understanding and executing the project's strategic approach to data management in data spaces:

- Theme 1: Data lifecycles in DS2
- Theme 2: Regulatory enablers relating to DS2
- Theme 3: Business cases and business models of the DS2 participants
- Theme 4: Data products in DS2
- Theme 5: Implementing enabling technology and DS2 modules

6.3.1 Theme 1: Data lifecycles in DS2

Modelling data lifecycles in DS2 will be based on the following four lifecycle layers as expressed in Figure 4.1.2.3.1:

1. Data lifecycle
2. Data product lifecycle
3. Data contract lifecycle
4. Data transfer lifecycle

This layered structure allows us to address both the automated opportunities to be embedded in the technical data space implementations, i.e. DS2 modules, and informational and transformational opportunities through regulatory and business-related aspects of data sharing in data spaces.

The first layer considers data lifecycle as a whole, extending from the boundaries of the technological data space building blocks to regulatory landscape evolution and business case considerations of individual data space participants for cross-data-space data sharing.

On the second layer, based on the business case aspects, data space participants need to specify their data products and express them as data product offers. These data products will be made public/available to other data space participants, and in cross-data-space cases to the participants of the other data spaces through a catalogue. On this second layer, e.g., visibility control becomes essential.

The third layer of data lifecycle starts from the achievement of a data product contract and ends with the termination of the same. It should be noted that on this level, not a single data product contract will be concluded, unless there is a mutual consensus between a data product provider and a data product consumer/user on the terms of such contract (e.g. quality, duration, restrictions of use, price). Such terms of contract highlight the usage control aspects, often as behavioural restrictions to be implemented outside the technological implementations.

On the final, fourth layer, the actual data transfer takes place, starting from providing the access to data and ending with the removal of the access. On this layer the technological implementations enabling access control emerge.

Assessment of the above data lifecycle layers from the perspective of cross-data-space data sharing will be based on the challenges faced on each layer. Based on the initial identification, for example the challenges identified below emerge on different layers. These challenges express the core focus that the data spaces aiming for cross-data-space data sharing should provide answers to.

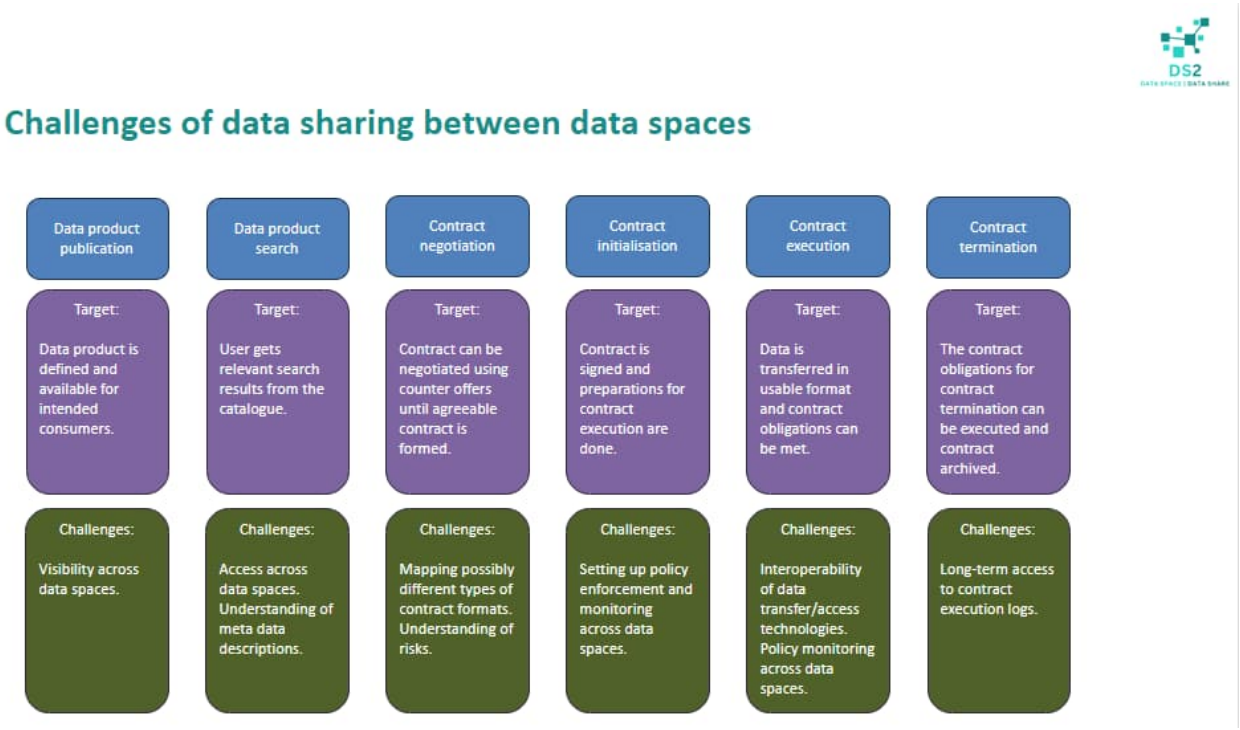

Figure 12 Challenges of Data Sharing

The initial mapping of the DS2 modules to the data lifecycles has been presented in D2.2 – Requirements, Baselines, KPI'S, Architecture and Specifications v1.0 – as follows:

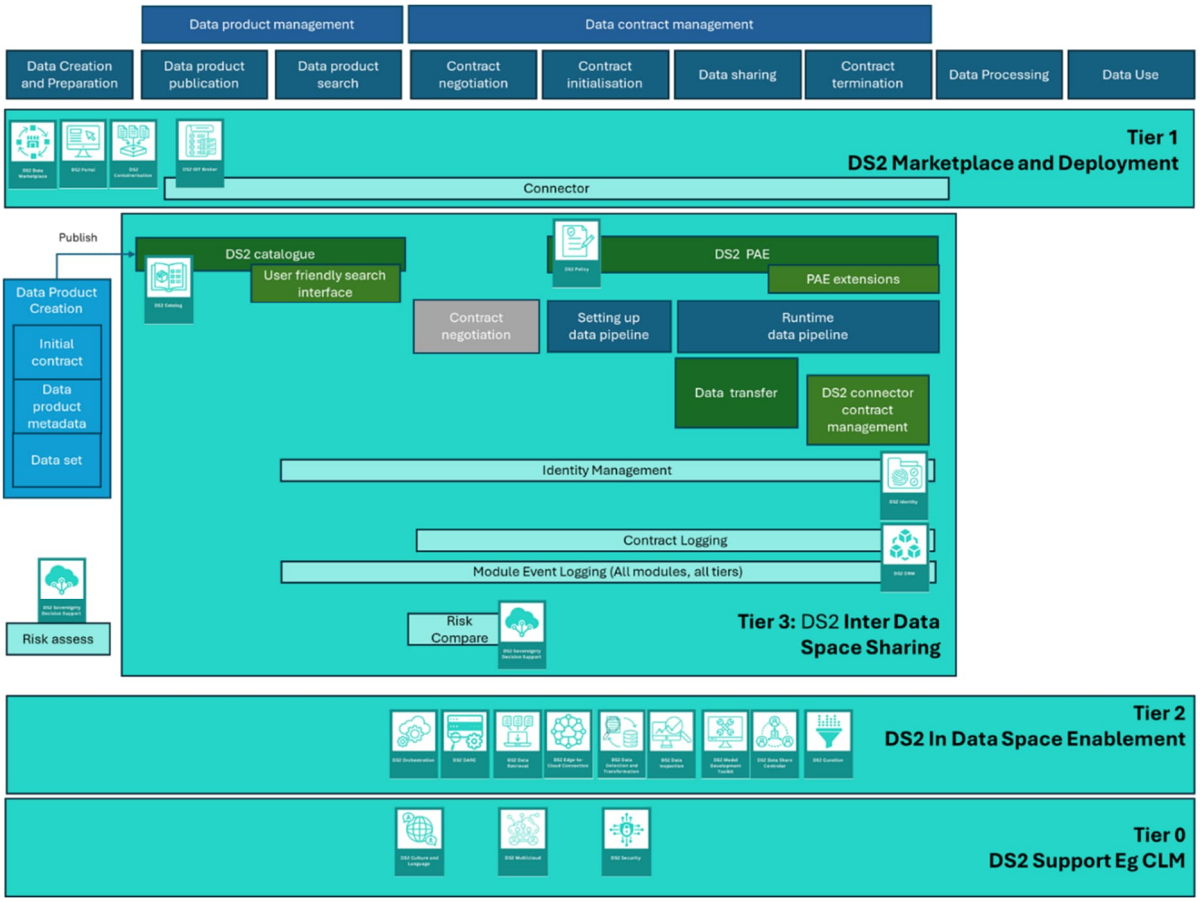

Figure 13 Mapping of the DS2 modules to the data lifecycles

Future needs:

During the next phase of WP3, it is needed to re-assess the data lifecycle layers and identify steps and characteristics of them, e.g., by identifying the focal DS2 modules to be used in the DS2 use cases and the core data lifecycles within the DS2 use cases. The assessment should also be broadened to cover cross-data-space challenges. The work should be done in close collaboration both with the use cases and the DS2 module providers.

6.3.2 Theme 2: Regulatory enablers relating to DS2

To assess the opportunities enabled by regulation, the following issues need to be identified:

- Mechanisms introduced by the regulations
- Anticipated data sharing roles of the participants
- Interactions taking place between the participants
- Type of data shared between the participants
- Types of intermediaries involved
- Types of services enabled

Based on these attributes, the main opportunities afforded by the regulation can be identified. On a general level, this approach is based on the concept of triggers presented in the DSSC Blueprint. DSSC classifies in its Regulatory Compliance building block the triggers into three categories: types of data, types of data space participants, and types of use cases (DSSC, 2024). We detail and develop this approach further by focusing on the underlying mechanisms of the regulations, the role the data space participants take in their interactions and types of actors involved in the use cases.

The first regulations assessed were the DGA and DA. These regulations entail elements that are enabling in their nature, and, thus, may cause disruptive opportunities. From the DGA and the DA it is possible to monitor these future opportunities, while, e.g., the DMA and the DSA aim to a higher extent regulate and set boundaries for existing platforms.

With regard to the DGA and the DA, we have studied the text of the regulations and their recitals, the guidance given by the regulatory authorities (European Commission, 2024b) and (European Commission, 2024a), and the studies about data intermediaries (European Commission. Joint Research Centre., 2023) and (Bobev *et al.*, 2023).

Based on the initial assessment, the following core regulatory enablers have been identified (see forthcoming ISPIM Osaka Conference Proceedings on "How the European Data Regulation Enables Innovation for Platform Ecosystems?", Suksi and Pussinen, 2024):

- DGA:
    - DGA art 2(11) and 12(a) require neutrality and separation of the ecosystem level intermediaries. This leads to opportunities for new types of intermediaries building on more inclusive forms of data governance, trust and data sovereignty.
- DA:
    - DA art 2(14), 5, 8 and 9 leads to decoupling of the services relating to the data of connected products. This opens opportunities for creating data-based services that for instance aggregate, enrich or transform the data.
    - DA art 2(12), 2(15), 2(16), 4 enables wider access to data for the users of connected products. This creates opportunities for intermediaries and service providers aiming to increase access to data, e.g. service providers acting as agents of data subjects.

Future needs:
The assessment of the Regulatory enablers should be continued through further regulatory analysis of the remaining regulations identified to be in the DS scope (see Section 3 above) and co-creation with the DS2 use cases. In the regulatory analysis, focus should be on the cross-data-space regulatory opportunities. Working with the use cases in an iterative manner allows to further scope the relevant regulatory analysis to those opportunities that are most potential to the use cases. This work includes the mapping of the roles of the use case actors to the roles recognised by the regulations and mapping the interactions taking place between them to the mechanisms introduced by the regulations. Finally, the cross-data-space opportunities identified from the regulation, should be mapped to Theme 1 on data lifecycles.

6.3.3 Theme 3: Business cases and business models of the DS2 participants

As stated in the DSSC Blueprint:

"The Data Space Value is the aggregate value generated from all data transactions and use cases within the data space. This value emerges from the collaborative efforts of the participants as they utilize the data space. Importantly, the definition of data space value is neutral regarding how this value is shared or captured among participants."

Therefore, it's a part of the participants business model to be aligned towards a concept of an open business model. An open business model concept grasps the fact that an organization accepts that some parts of their business model remain dependent on other parties for their contribution (or data) (Frankenberger, Weiblen and Gassmann, 2014). To develop organizational business model from a closed business model to an open business model, organization needs to engage in business model innovation. Business model innovation refers to the organization's capability and visionary to either *transform* their existing business models as a response to outside influences and technical development, or *create* a new business model based on an identified value creation opportunity (Ahokangas and Myllykoski, 2014).

In a data space setting, participants participate in value co-creation in a business ecosystem like setting, where the ecosystem is formed from potentially multiple different data spaces. Each partner can be stated to participate in the co-creation of different value creation dimensions, mainly value co-creation, co-delivery, and co-capture (Amit and Zott, 2001). It is understood that these same value creation dimensions are shared with both potential business cases as well as business models. The data sharing use cases may well generate multiple business cases for data sharing.

Use case development begins with the identification of potentially valuable data sharing scenarios between different actors. These scenarios can be implemented in data spaces or without data spaces relying on 1-on-1 contracts. Based on the DSSC blueprint: "*A data space use case is a specific setting in which two or more participants use a data space to create business, societal or environmental value from data sharing."* With multiple actors, the use of data space technology for the implementation of use cases is particularly beneficial when numerous participants engage in data sharing (making point-to-point integrations challenging and emphasising the issues of data sovereignty and trust) or when the same data can be utilised across multiple use cases (DSSC, 2024).

For the business models and related business cases for data sharing to be utilized, the data needs to be in an interoperable form. This is where the definition of a data product comes into play. The definition of data products is still evolving in the data space community. In the DSSC blueprint, a data product is defined *as a data sharing unit, bundling resources (data and/or data services) and metadata that describes the license terms, the resources, and other information in a machine-readable way* (DSSC, 2024). While the content of data products may vary, we talk of a package of resources and all information that a potential data product consumer needs, to make a decision whether to consume the data product or not.

Future needs:

In the next phase, the use cases of DS2 should be examined from the value generation perspective to outline the potential business cases of data sharing between data spaces. The potential value created in the use cases may be financial value, environmental or societal value. The role of data products in the value creation of the use cases should also be analysed.

6.3.4 Theme 4: Data products in DS 2

Detailed information on Data products in the data spaces is included in the DSSC Blueprint.

It identifies a data product as:

> *Data product is a data sharing unit, bundling resources (data and/or data services) and metadata that describes the license terms, the resources, and other information in a machine-readable way. While the content of data products may vary, we talk of a package of resources and all information that a potential data product consumer needs, to make a decision whether to consume the data product or not.*

A data product typically includes

- *the resource, which can be data and/or data service,*
- *the description of the resource,*
- *its allowed purposes of use,*
- *quality, format, frequency, duration and other requirements the data product fulfils,*
- *access and control rights (e.g., attribution, Intellectual Property Rights, liabilities, geographical limitations, usability for training LLMs),*
- *delivery options (e.g., APIs, SMTP, web interface, mapping tools),*
- *information about data provenance and lineage,*
- *pricing and billing information,*
- *other information (e.g., ethical considerations),*
- *and metadata describing all these above. (DSSC, 2024)*

It also explains why data products are needed:

> *Perceiving data as a product will, for many organizations, require a change of mindset – the idea of producing data with reuse in mind is not commonplace yet.*
>
> *From the data product owner's perspective, it is beneficial that data products containing the same data can be delivered to multiple use cases. Therefore, it is recommended that data product owners consider to develop multiple data products with various options of the associated information (e.g., different license terms, delivery options, commercial and technical requirements).* (DSSC, 2024)

The *DSSC* Blueprint contains also a conceptual graph of the elements and concepts involved in building a data product:

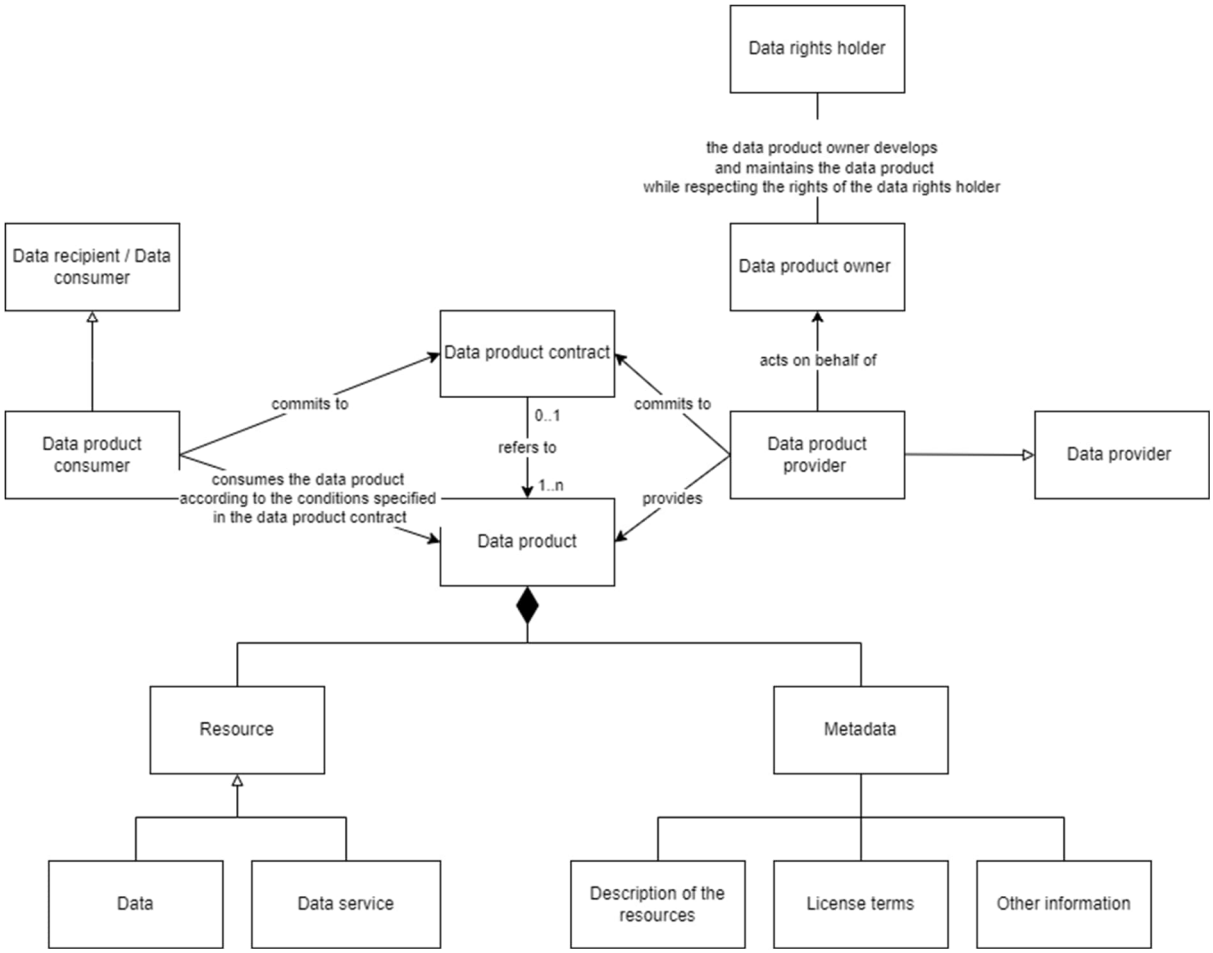

Figure 14 1 Data products and related concepts (DSSC, 2024)

In addition, the CEN CENELEC Trusted Data Transactions Working Group definition of data products as "data sharing units, packaging data and metadata, and any associated license terms" (DSSC, 2024), (CEN-CENELEC, no date) should also be kept in mind.

The above concept of the Data product is useful especially as an interlinking building block between business and legal aspects of the use cases and the technological enablers to be implemented. The upper part of the graph identifies the roles of different actors in a data transaction and the bottom part the resources transacted on and means of communicating on the resources.

When looking at Data products from a theoretical perspective, it can be seen as a vehicle to address the loopholes and opportunities (see Section 4.1.2). For example, as summarized in Figure 5.2.3. below, there can be data rights holders, data product owners, data providers, data consumers, and data recipients who can have different beliefs, which can lead to them seeing the same regulations and requirements differently.

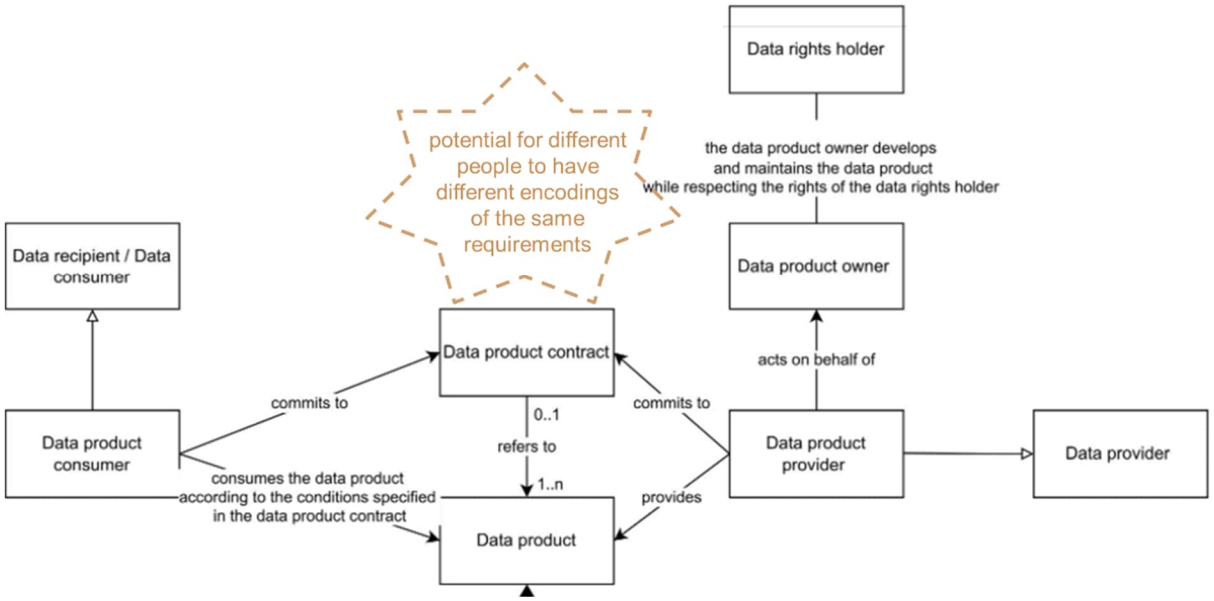

Figure 15 Different people who can have different beliefs about the same requirements

As summarized in Figure 5.2.4.3, techniques identified in Section 4.1.2 can provide supporting structures for the preparation of data product contracts, which otherwise could be seen as being incomplete from the view of some participants [PP,QQ]. Such supporting structures may include methods like the Opportunity-based data lifecycle management or risk-based tools like the Spyderisk-tool.

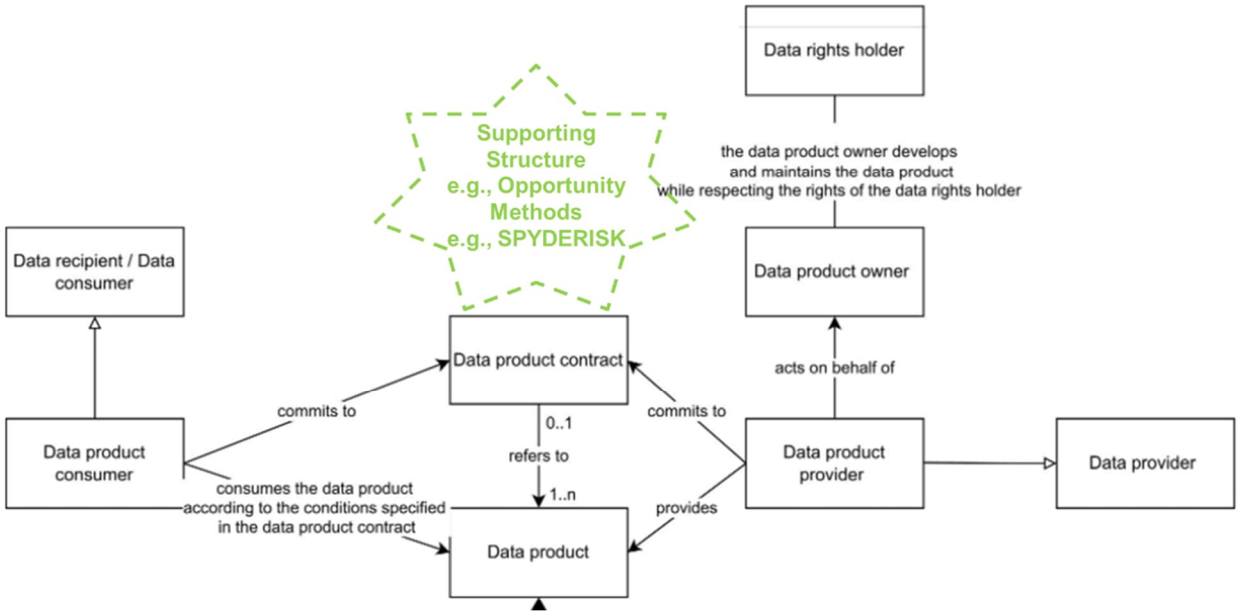

Figure 16 Support structure for the preparation of data product contracts

The current status of specification of the Data products in DS2 is that the first workshop with the use cases and DS2 module providers has been held. The Data product concept has been explored and applying it to the DS2 use cases has been initiated through a template to assist in identifying the use case specific data products

and an end-to-end mapping concentrating on the technological aspects of the Data product (i.e. Resources a) data as an asset with data types, data format and b) data service with interface and access information).

Future needs:
The work initiated in specifying the data products for the DS2 use cases should continue. The Data products need to be co-created both with the DS2 use case participants and the DS2 module providers. The results of Theme 2 (Regulatory enablers) and Theme 3 (Business cases) should be fed into the identification of Data products.

6.3.5 Theme 5: Implementing enabling technology and DS2 modules

Development of the business cases and identification of open business models specify the use cases in such a manner that the underlying regulatory environment, technological capabilities and business opportunities can be considered from the perspective of all data space participants involved in data sharing. Once business case development and business model innovation are mature enough and corresponding data product specifications have been made, the Data product needs to be aligned to the enabling data space technology. In this respect, we have identified the first major implementation example aiming to interlink different data life-cycle layers (see Section 5.2.1) and identify those actions that can be automated by data space technology and DS 2 modules and those actions that will need to be handled by supporting structures with human involvement. This example is laid down below and it addresses the question "How to operationalize business driven licensing into the data spaces policies?"

This step will identify practical measures that can be taken to utilize the enabling data space technology to take advantage of the opportunities identified in the previous steps. These measures should be taken after the specification of a data product has been made – including the identification of the underlying business case.

6.3.5.1 Implementation example: From business cases to policies

When considering the transformation of the business cases into machine-readable policies, two steps should be taken. Firstly, the scope of use of the data space participants should be aligned with the business cases developed. This scope of use is often described as licenses or usage rights. Secondly, once the business perspectives are coded into legal terms, the licenses may be encoded into technically implemented policies. The first step involves a link to the DSSC building block Contractual framework and the second step to the DSSC building blocks under the Data sovereignty and trust pillar.

6.3.5.1.1 Background on smart contracting

Not all license terms can or should be automated. What should and what should not be encoded into technology depends on the business cases and business models of the data space participants and available technological capabilities.

Smart contracting is a term that relates to the application of digital technologies and artificial intelligence technologies to contracts, e.g., licenses. Technologically smart contracts can include the preparation, recording, and administration of contracts being digitalized, machine-readable, and self-executing. At a low level of technological smartness, contracts can involve adding electronic signatures to digital versions of contracts that have been composed through traditional practices. At a higher level of technological smartness, contracts could be prepared by and executed by artificial intelligence (AI).

Considering for instance AI within smart contracting, an underlying rationale for the application of AI in contracts is that legal regulations are a type of algorithm that has traditionally been described in natural language but can also be described in machine-readable language. Algorithms can be described as "A finite set of unambiguous instructions that, given some set of initial conditions, can be performed in a prescribed sequence to achieve a certain goal and that has a recognizable set of end conditions" (American Heritage, no date). However, such definitions of algorithms draw attention to the fundamental challenge for contracts, both traditional and technologically smart, which is to eliminate all ambiguity (Solan, 2004; Koc and Gurgun, 2022). As illustrated by reports with titles such as, "Hallucinating Law: Legal mistakes with large language models are pervasive" (Dahl *et al.*, 2024), there are fundamental limitations to the potential of AI to reduce ambiguities in the preparation and administration of legal contracts. For example, there can be deteriorations in AI performance, which may be referred to as "performance drift" (Roschewitz *et al.*, 2024), "model amnesia" (Shumailov *et al.*, 2024), and "model collapse" (Fox, 2024). Such fundamental issues with AI add to the established problem of "Garbage In Garbage Out" by adding AI generation of garbage to whatever human errors might be inputted into a contract. Thus, technologically smart contracts have potential to be colloquially-speaking stupid contracts.

Overall, as summarized in Figure 17, the introduction of digitalization or artificial intelligence into the preparation and administration of contracts can increase the extent of entropy and complexity to be managed. The higher the entropy and complexity to be managed, the more potential there is for information uncertainty, physical disorder, and unproductive energy expenditure. Importantly, all living things have evolved to try to manage entropy and complexity with least action (Fox, 2024). The principle of least action is expressed in relation to information management in the principle of least effort (Chang, 2016) and the principle of least collaborative effort (Davies, 2007).

Accordingly, it is important that the formulation of technologically smart contracts relating to data spaces should be focused on achieving unambiguous minimal contract descriptions with least action. The more technologies are involved, which are based on different underlying ontologies and different languages, the more difficult it will be achieve unambiguous minimal contract descriptions with least effort. Hence, there should be selective targeted applications of technologies for use in technologically smart data product contracts. In particular, technologically smart data product contracts that have the elegance (Glynn, 2010) of unambiguous minimal description should be preferred to the clunky complicatedness of ambiguous contracts that have descriptions in multiple languages.

6.3.5.1.2 License terms and policy categories from existing data sharing initiatives

One of the core elements in the Data product concept is the metadata relating to the license terms. When aiming to drive opportunity-based business in the context of cross-data-space data sharing, in the first phase, this means identifying the core license types that enable seizing such opportunities, and in the second phase selecting those license types that take best advantage of cross-data-space related business opportunities. Additionally, these opportunities should be such that can utilize the data space technological enablers, and make the license elements machine readable and executable, i.e. building smart contracting into the licenses.

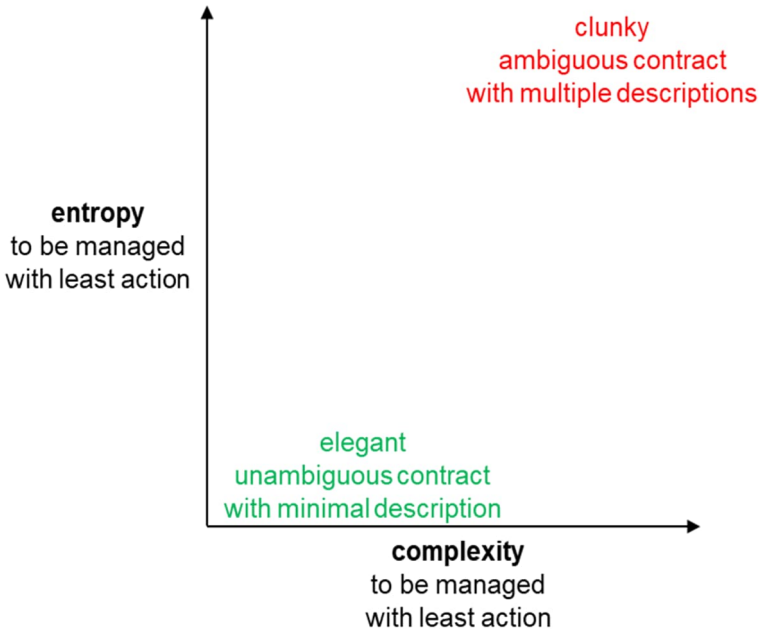

Figure 17 Elegant contracts have low entropy and low complexity

To advance these goals, the DS2 team has identified a variety of license scopes and restrictions from available data sharing initiatives and rulebook-templates. Most of these are not directly addressing data spaces, the concept of which is still evolving. However, these are some of those initiatives that form the basis for the development of data spaces.

iShare Trust Framework

The iShare trust framework encompasses a license code list of 9 licenses varying from non-limited to fully determined between the parties (Licenses | iSHARE Trust Framework). In the iShare Trust Framework these are identified as "instructions" to which the parties are bound by the underlying contract and scheme rules. A categorization of the license codes in the iShare trust framework (iShare Foundation, 2024) can be as follows:

- Unlimited use allowed:
  - license nr. 0000 (no limitations)
- Re-sharing restrictions:
  - license nr. 0001 (resharing within contracting parties)
  - license nr. 0002 (internal use)
- Limited type of use:
  - license nr. 0003 (non-commercial use)
- Enrichment restrictions:

- license nr. 0004 (only with own data)
- license nr. 0005 (other parties's data allowed)

- Combined restrictions on enrichment and re-sharing)
  - license nr. 0006 (only with own data + non-commercial basis re-sharing)
  - license nr. 0007 (other parties's data allowed + non-commercial basis re-sharing)
- Fully negotiated between the parties (license nr. 9999)

Data sharing canvas

The Data sharing canvas of the Data sharing coalition identifies initial policy clusters and polices. The collection is non-exhaustive example meant to help the development of cross-domain data sharing. It contains 17 clusters and 43 policies based on their data sharing initiatives (data-sharing-canvas.pdf).

The policies are divided into two main types: "Access control rules" and "Obligations and advice". The latter includes both obligations relating to usage and other types of obligations.- (Data Sharing Coalition, 2021)

Clusters under "Access control rules":

- Authorisation
- Authentication
- Privacy (pre)
- Information classification
- Information access
- Laws and regulations
- Employee qualifications

Clusters under "Obligations and advice":

- Scope
- Liability
- Privacy (post)
- Operational conditions
- Provenance

- Data storage
- Non-repudiation (digital signature requirement)
- Information security
- Geographical information
- Supervision (all)

Sitra Rulebook for a fair data economy

In the Sitra Rulebook, the license terms under which a data provider grants a right to use the data is to be described in the "Dataset Terms of Use". The Dataset Terms of Use leaves the license terms open to all kind of deviations and are merely placeholders for considering all relevant aspects. The users should consider preparing more specific templates to reflect the business context.

The basic default set of license terms in the Sitra Rulebook (*Rulebook for a fair data economy, version 2.0*, 2022) regarding usage scope or restrictions contain:

- Core restrictions:
    - The license is non-exclusive
    - Data provider may seize the provision of data by [30 days] notification
    - Data provider may change the terms by 90 days notification
    - The license to use already received data will remain in force despite termination
- Redistribution of data:
    - Default: Redistribution allowed to other parties (and their affiliates)
    - Optional: Redistribution allowed to third party end users
    - Restrictions on processing and redistribution of data: [open list, to be listed case by case]
- Core borderline between data and derived material:
    - Default: Restrictions relating to data do not apply to derived material and redistribution of derived material allowed for any third party
    - Default: The borderline between data and derived material is specifically defined to exclude misuse with minor modifications of data.

- Restrictions on the use and redistribution of derived material: [open list, to be listed case by case]

- Additional obligations:
  - License may be subject to a payment obligation (license fee)
  - License may be subject to reporting obligations
  - License may be subject to audit obligations
  - License may be subject to data security obligations
- Carve-outs:
  - Use of learnings, added skills and experience not restricted
  - Robotic process automation, machine learning and is allowed
- Liability borderline:
  - Liability for possessing the right to make data available is on the data provider
  - Liability regarding the use of data (only as-is, as available, no warranty) is on the data consumer/data user
- Default regimes apply to:
  - Confidential information included in data may be subject to restrictions
    - Default: Restricted to trade secrets and clear markings
    - May be subject to specific terms
  - Personal data included in data
    - Subject to data protection legislation, including GDPR
    - May be subject to specific terms
  - Intellectual property rights:
    - Default: No transfer of intellectual property rights
    - May be subject to specific terms
- Any additional restriction allowed

From the DS2 use cases, Green Deal includes the DIH AGRIFOOD Data Space (DADS) that has built their governance model based on the Sitra Rulebook. In terms of licenses, DADS has developed further the Dataset terms of use by categorising useful license terms as follows:

Option A: "Open data" terms and conditions

Option B: ''Restricted/closed'' terms and conditions:

B1: Restrictions in terms of User

1. All users have usage rights
2. Only identified Users have usage rights: (list of Users), the rest of Users will be granted rights as per specific Request for Data
3. All Users will be granted rights as per specific Request for Data

B2: Fees and Payments

1. The use of Data is not a subject to fees and charges.
2. The use of Data is subject to following fees and charges: [text]

B3: Personal Data

1. The Data includes no personal data and/or the Data Owner is giving a permission to the User to receive and process personal data.
2. The Data includes personal data, and its reception and processing are subject to the following restrictions: [text]

B4: Other restrictions and/or limitations

1. The Dataset includes Confidential Information and consequently its use and processing are subject to: [text] or NOT APPLICABLE
2. The use of Data is subject to the following specific data security obligation: [text] or NOT APPLICABLE
3. The use of Data is subject to following specific Intellectual Property Rights obligation: [text] or NOT APPLICABLE
4. The use of Data is subject to the following specific audit obligations: [text] or NOT APPLICABLE
5. The use of Data is subject to the following duration (time) limitations: [text] or NOT APPLICABLE
6. Is the use of Data subject to any other restrictions or limitations? [text] or NO

IDSA Rulebook

The IDS contract framework builds upon the SITRA rulebook and additional components for contract terms are envisioned (International Data Spaces Association, 2024):

> *Therefore, the IDS contract framework will focus on additional components and guidance highlighted in different use cases of data sharing implemented under the IDS specifications. These may include domain-specific dataset terms of use templates or more detailed components for cross-continent data sharing or privacy. If the IDS contract framework requires modifications to the SITRA rulebook's terms and conditions, they will be proposed also to Sitra's workgroup to maintain compatibility and to avoid different versions of terms and conditions.*

6.3.5.1.3 Data policies in the DSSC Blueprint 1.5

DSSC Blueprint recognizes only two types of data policies. The "Data access policies" are aimed at controlling the access to data and the "Data usage policies" are aimed to restrict the usage of shared data (DSSC, 2024).

Terminology relating to policies has been described in the DSSC Glossary (DSSC, 2024) as follows:

Data policy

A set of rules, working instructions, preferences and other guidance to ensure that data is obtained, stored, retrieved, and manipulated consistently with the standards set by the governance framework and/or data rights holders.

Explanatory text: Data policies govern aspects of data management within or between data spaces, such as access, usage, security*, and hosting.*

Data access policy

A specific data policy defined by the data rights holder for accessing their shared data in a data space.

Explanatory text: A data access policy that provides operational guidance to a data provider for deciding whether to process or reject a request for providing access to specific data. Data access policies are created and maintained by the data rights holders.

Data usage policy

A specific data policy defined by the data rights holder for the usage of their data shared in a data space.

Explanatory text: Data usage policy regulates the permissible actions and behaviours related to the utilisation of the accessed data, which means keeping control of data even after the items have left the trust boundaries of the data provider.

In the DSSC Building Blocks the role of the policies is restricted to technical building blocks under the pillar of "Data sovereignty and trust". Therefore, the policy guidance in the DSSC Building Blocks is restricted only to "Access & usage policies enforcement", which supports the data space participants to implement and enforce access and usage policies through their systems and applications. It includes the following capabilities (DSSC, 2024):

Policy Definition and Governance

- *Define policies specific to the data space's rulebook.*
- *Set policies by the data rights holder.*

Policy Implementation and Enforcement

- *Consolidate policies and rules into a machine-readable and executable format.*
- *Implement mechanisms for obtaining and managing auditable consent for data usage, applicable to all types of data where consent is required.*
- *Supports policy negotiation among Data Space participant, using an engine.*
- *Enforce the execution of the agreed policy.*

Services for implementing technical building blocks include for instance "Policy Information Point Services" under the "Federation services" distinguishing the following categories (DSSC, 2024):

- *Identity provisioning*

providing information on the identity of a person or asset.

- *Personal Consent*

indicating whether a person has given consent for sharing data. This is particularly important when consent is necessary based on privacy legislation (e.g. GDPR). The personal consent service can be deployed (as is the case for all federation services) in different ways. For example, there can be a single service for personal consent in the data space, but there can also be multiple services.

- *Conformity Assessment*

assessing whether someone is conforming to a specific policy. This could be relevant, e.g. when onboarding a new participant. A set of credentials might need to be supplied to assess conformity before a participant's credential can be issued.

- *Data Usage Policy Issuing*

issuing a policy (e.g., a standardized ODRL policy) that can be used by participants of the data space (as an alternative to a policy defined by the individual participant).

The policy information points (PIPs) connect to policy decision points (PDPs) and policy execution points (PEPs) in the participant agent.

It should be noted that the DSSC Blueprint does not directly refer to the category of Visibility controls. The "Publication and discovery" building block sets as a pre-condition that the data product consumer is a registered data space participant. When considering broad and confidential data sharing, the issue of restricting the visibility of metadata emerges and is further addressed below.

6.3.5.1.4 DS2 Classification of policies

Policies can be either expanding or limiting the use of the data products. In other words, policies deny access unless conditions are met, or policies allow access unless conditions limit the access. As stated above, policies can be categorised to 3 classes: access policies, usage policies and visibility control. As stated above, DSSC defines only access and usage policies, visibility policies being prospect in the future. Further, the usage policies can be applied in pre-contract and post-contract period. Both will be discussed below, and focus is on pre-contract usage policies.

Access policies

According to DSSC, access policies are merely internal operational guidance policies for data provider by the data rights holders to eventually create usage policies. They are not to be implemented by technical means, but one of their purposes is to act as business requirements for the usage policies. Business requirements may for example define options on how the data products are being offered to the data consumers. Such rules may include possibility of time constraints, individual licensing, or company licensing etc. The access policies can also guide personnel on managing the data that is collected from the data rights holders during operations. As an example, data protection or data privacy consents can be included in the operational procedures. Access policies should also guide company on how to enforce access control actions originated outside the organisation. Such case occurs for example if an individual controls access to their personal information.

Usage policies

Pre-contract usage policies and often referred to as access control, which is the traditional way of technically limiting the access to the resource (data product). Data spaces expand access control policies to allow providers to create access conditions based on certain parameters. These parameters are listed in IDSA web page (https://w3id.org/idsa/code)(*International Data Spaces Information Model*, no date). The parameters also include other actions, such as logging events or notifying e.g. clearing house on the transactions. Some examples of advanced (not only allowing or denying access) access control policies are listed below:

- Allow access only from certain connector
- Time duration for access (e.g. 24 hours)
- Time interval for access (e.g. between two dates)

```
{
  "@context": {
    "xsd": "http://www.w3.org/2001/XMLSchema#",
    "ids": "https://w3id.org/idsa/core/",
    "idsc": "https://w3id.org/idsa/code/"
  },
  "@type": "ids:Permission",
  "@id": "https://w3id.org/idsa/autogen/permission/f25e6618-14d7-474c-aefc-d81049419786",
  "ids:description": [
    {
      "@value": "",
      "@type": "http://www.w3.org/2001/XMLSchema#string"
    }
  ],
  "ids:title": [
    {
      "@value": "",
      "@type": "http://www.w3.org/2001/XMLSchema#string"
    }
  ],
  "ids:action": [
    {
      "@id": "https://w3id.org/idsa/code/USE"
    }
  ],
  "ids:constraint": [
    {
      "@type": "ids:Constraint",
      "@id": "https://w3id.org/idsa/autogen/constraint/9e8b165b-d592-4e95-86d2-b197ffeb5a61",
      "ids:operator": {
        "@id": "https://w3id.org/idsa/code/LTEQ"
      },
      "ids:leftOperand": {
        "@id": "https://w3id.org/idsa/code/COUNT"
      },
      "ids:rightOperand": {
        "@value": "2",
        "@type": "xsd:double"
      }
    }
  ]
}
```


Figure 18 Example of a usage policy

- Access count (e.g. user can download the resource 10 times)

Usage policies are technically implemented using standard Open Digital Rights Language (ODRL). Figure above is an example of access control policy that permits access to the resource, but only in the constraints are met. The constraints include a count with value of 2, meaning that the resource can be accessed only twice, and after that the access will be denied. The benefit of using ODRL is that it defines the policies unambiguously, which is of course needed in machine-to-machine communications. The policies must be visible and understandable to the end users as well, and this will require an ODRL translator to natural language. In simple terms usage policies controlling access to the resource can be understood as a series of conditions, which all must pass to get access to the resource. In case any of the conditions (policies) provide a negative response, the overall response is also negative (i.e. no access).

Sometimes providers would like to also control usage after the contract has been signed. As data products are non-rival, the usage control is often hard to achieve by technological means. If data consumer has a data product in their possession, technically it is difficult to limit how it is used. There are some mechanisms for

that (one being that the data is not shared in the first place, just the needed results, but then it can be discussed what was the actual shared data). Such usage policies are most often defined by licensing and not by technical means. In this case the enforcement is reactive and done only after the policy breach has happened. Of course, the mere existence of the policy reduces the breaches: most people in normal operations do not breach the license terms intentionally. Usage control is one of the main tools for any business to support the selected business model on data sharing. When the data has been shared, it is typically technically out of control of the organisation and defending rights of that data can be managed with licensing.

Visibility control

Data space defines a data catalogue which displays metadata of the data products in the data space. Such data catalogue can be centralised (as in IDSA architecture), where all connectors advertise their data products' metadata in the joint catalogue. Data catalogue can also be decentralised (as allowed by EDC architecture), where a data space search initiates queries in all connectors of the data space. Visibility control means that data provider can limit the visibility of data products (i.e. metadata of the data products) using visibility policies. Examples of such visibility policies would be limiting visibility by user, connector (from which the query was originated) or geolocation. Visibility policy can be seen as more restrictive than access policy as the restricted users not only don't have access to the resource, but they are not allowed to even see the mere existence of the resource. IDSA architecture currently does not allow visibility control by specification. All data products in the data catalogue are visible to all users in the data space. Of course, providers can limit access to the resource, but they cannot thus hide resources from anyone. Visibility control can be used by organisations that want to share data with some of the data space participants, but the data is either secret or it is customized to single use. In such cases organisation might choose to limit the visibility from other users. This can protect own business (hiding what organisation has) and/or minimise unnecessary requests when organisation is certain they will not grant access to requestor anyway.

6.3.5.1.5 Table of policies

As examples some different policies were created in ODRL syntax. The policies can either allow permission and limit it with parameters or prohibit permission and make exceptions through parameters. The table below illustrates some of the example policies and the meaning of these policies in natural language.

	Type	Action	Duty	Duty value	Constraint base 1	Constraint value 1	Constraint base 2	Constraint value 2	Translation
ConnectorRestrictedUsage.jsonld	Permission	Use			Operator	https://www.google.com			Access if your domain is google.com
DurationUsage(24h).jsonld	Permission	Use			Elapsed time	PT24H			Access for 24h from creation
N(10)TimesUsage.jsonld	Permission	Use			Count	2			Can access twice
ProhibitAccess.jsonld	Prohibition	Use							Access prohibited
ProvideAccess.jsonld	Permission	Use							Access provided
UsageDurationInterval.jsonld	Permission	Use			After	2024-08-29T00:00:00Z	Before	2024-08-31T00:00:00Z	Access between two timestamps
UsageLogging.jsonld	Permission	Use	Log						Access provided, all actions logged
UsageNotification.jsonld	Permission	Use	Notify						Access provided, a party (provider) is notified when accessed
UsageUntilDeletion.jsonld	Permission	Use	Delete	2024-09-01T00:00:00Z	After	2024-08-29T00:00:00Z	Before	2024-08-31T00:00:00Z	Access between two timestamps, and delete automatically

Table 3 Example policies and the meaning of them

It should be noted that the fields in the table above are not exclusive. IDSA specification allows a lot of other attributes to the policies. As stated in the previous chapter, ODRL creates a readable structure using certain

fields that are unambiguously interpreted. Equally it is clear that users do need to understand the policies, and for average end user ODRL might not be the best way of communication. Therefore, translation to plain English will be needed. It would also be beneficial to be able to translate plain English into ODRL automatically, to help users to create policies they want to apply to their data products.

6.3.5.2 Future needs

Further implementation examples should be considered in the next phases of DS2. These may be developed around the other elements of the Data product concept, e.g., data rights management or value-creation services.

Regarding the first example on how to operationalize business driven licensing into the data spaces policies, the usefulness of the collection of license terms and policies should be assessed from the perspective of the DS2 use cases and identify those that can have cross-data-space relevance or impact. Available licensing assistants and license generators should also be investigated.

6.4 Enforcing policy and contract management

DS2 uses the IDSA specification as the foundation for defining how data sharing is performed among dataspaces. Specifically, DS2 uses the EDC Connector as a technological component that follows IDSA specifications. Consequently, the IDSA specification will guide the requirements for the Policy Enforcement module in DS2, while the EDC Connector will dictate how those requirements are implemented in software.

It is important to highlight that policy enforcement requires trusted sources of metadata about both the participants and the shared data.

6.4.1 Functional Requirements

DS2 uses policies to enable secure, trusted, and controlled data sharing between participants. These policies are agreed during the contract negotiation phase, serving as the mechanism to ensure that data is used in compliance with the terms established between the provider and the consumer.

The requirements for DS2 policies are:

- Policies must be linked to the assets being shared (e.g., data) to define who can access the data and how it can be used.
- Policies must define access rules to ensure that only authorized participants can access the data.
- Policies must specify allowed actions and the associated conditions or restrictions that govern how the data can be used after access is granted.
- Policies associated with assets must be accessible for consultation before the consumer begins the negotiation process.
- Policies must be written in ODRL (Open Digital Rights Language), the reference language adopted by IDSA, to ensure they can be interpreted and processed by software tools.
- The evaluation of policies must rely on metadata about participants and data, which is obtained from other modules within the DS2 ecosystem.

- The DS2 policy enforcement system must be extensible to support the definition of new metadata types and rules as requirements evolve.

6.4.2 DS2 Solution.

In DS2 the policy enforcement will take advantage of the capabilities of the PAE module that sits on top of EDC. DS2 will develop the required extensions for EDC Connector to achieve the functional requirements to ensure that both access and usage policies are enforced over the data sharing.

EDC Connector Extension

The EDC Connector comes with a Policy Engine capable of processing ODRL policies. However, it is the responsibility of DS2 to implement the specific behaviour for each type of policy. For example, the Policy Engine can evaluate a rule based on checking the value of an attribute. Yet, it is DS2's responsibility to provide the attribute value at runtime, which involves retrieving metadata from the appropriate sources. Additionally, DS2 must define and implement the actions that should be taken based on the result of the attribute evaluation.

```
{
 "@context": {
  "@vocab": "https://w3id.org/edc/v0.0.1/ns/"
 },
 "@id": "high-trust-policy",
 "policy": {
  "@context": "http://www.w3.org/ns/odrl.jsonld",
  "@type": "Set",
  "permission": [
   {
    "action": "use",
    "constraint": {
     "@type": "AtomicConstraint",
     "leftOperand": "trustLevel",
     "operator": {
      "@id": "odrl:eq"
     },
     "rightOperand": "high"
    }
   }
  ],
  "prohibition": [],
  "obligation": []
 }
}
```

Figure 19 Policy snippet example

For example, in a simple case where the goal is to validate if an attribute equals a specific value, DS2 must specify the actions for both possible outcomes (i.e., when the attribute value matches the expected value and when it does not). For example:

- Condition: check if the “trustLevel” attribute equals “high”
- Action:
  - If “trustLevel” == “high”, allow access
  - If “trustLevel” != “high”, deny access

Even this simple example can become more complex incorporating additional operators such as “greater than”, “included in a list”, etc.

For each attribute that requires validation, DS2 will provide the evaluation logic to support all the applicable operators for that attribute.

ODRL Policies

Policies will be defined using ODRL (Open Digital Rights Language[17], a standard used by IDSA and, consequently, by the EDC Connector It has the semantic to define both access and usage rules required in dataspaces.

An example of simple policy can be seen in the next block of code:

The exemplar policy specifies that for the action "use," the "trustLevel" must be set to "high." Once defined and stored, the policy can be assigned to any asset a provider wants to share. The combination of the asset and the policy constitutes the offer. If the consumer accepts the offer during contract negotiation, it results in a contract agreement.

It is important to note that creating the contract agreement is not the responsibility of the Policy Enforcement module. The role of Policy Enforcement is to ensure that the policies included in the agreement are enforced and satisfied.

Enforcement Process

The enforcement process will consist of several steps as shown in the next figure:

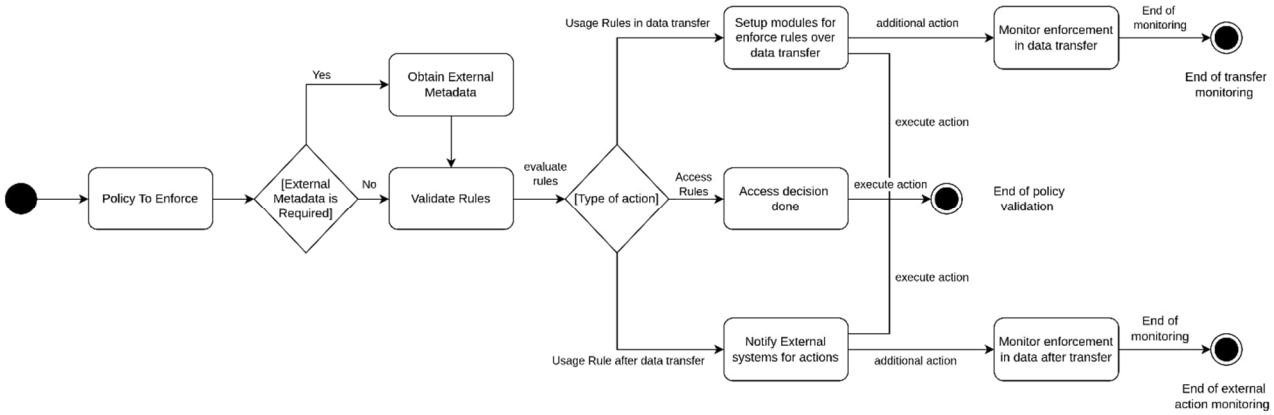

Figure 20 Enforcement Process

The process begins when a policy needs to be evaluated at a Policy Enforcement Point (PEP). The EDC Connector defines the available PEPs, so the process is triggered by the EDC Connector. Next, the DS2 extension determines whether additional metadata from external systems is required. If necessary, it obtains the metadata, for example, from the Catalog or Portal modules. With the metadata and the policy definition, the DS2 extension evaluates all the rules, executing the corresponding action for each rule. The rules will be of two main types:

- Access rules: the associated action will be allowed or not access to the asset.
- Usage rules: there will be two sub types:
    - Usage rules within the transfer process. For example, remove a column in the exchanged data.

[17] https://www.w3.org/TR/odrl-model/

- Usage rules after the transfer process. For example, the consumer must delete the data after n days.

The action for access rules is executed immediately in the evaluation process. The actions for the usage rules are delayed until the transfer process or after the transfer process. These rules make use of additional tools for executing the actual action. For example, making transformations to the shared during the data transfer and sending notifications to remind the consumer to delete data after a specified period.

The Policy Agreement and Enforcement module monitors the status of usage rules to ensure compliance. This includes tracking whether actions such as data deletion or usage limitations have been fulfilled.

Tracking

All decisions and actions undertaken within the policy enforcement process will be logged in the DRM system to ensure a comprehensive and traceable record. This logging mechanism will capture critical details, including timestamps, user information, and the nature of the actions performed. By maintaining this detailed audit trail, the system ensures transparency, supports accountability, and facilitates compliance with regulatory and organizational requirements.

6.5 Logging processes

Dataspace connectors are designed to facilitate secure and controlled data exchange between organizations by providing robust data rights management and logging capabilities. Data rights management allows data providers to define and enforce usage policies, ensuring that only authorized parties can access data under specific terms and conditions. This includes setting permissions, access controls, and usage restrictions to protect sensitive information. Additionally, logging actions create detailed records of all data transactions and access events, offering transparency and traceability. This comprehensive logging supports auditability and compliance with regulatory requirements, fostering trust among data exchange participants.

Connectors like the Eclipse Dataspace Connector (EDC) and those following the International Data Spaces Association (IDSA) standards handle logging and data rights management (DRM) through integrated technical solutions involving databases. These logs are often stored in databases like PostgreSQL or MongoDB to ensure persistence, auditability, and compliance with regulatory requirements. For DRM, these connectors employ policy languages like the Open Digital Rights Language (ODRL) to define data usage terms and conditions. The policies are stored in databases and enforced at runtime using policy evaluation engines that check incoming data requests against the stored policies, ensuring secure and compliant data exchanges.

Despite their robust features, the reliance on databases for storing logs, DRM and policies introduces potential single points of failure, scalability challenges when dealing with large volumes of data and necessitates robust backup and disaster recovery solutions to maintain data integrity and availability. Additionally, databases can be easily altered, leading to a loss of data integrity and compromising the reliability of the stored information.

A permissioned blockchain network offers a superior approach for storing logs, DRM, and policies due to its decentralized architecture, which eliminates single points of failure and enhances scalability. By distributing data across multiple nodes, it ensures high availability and fault tolerance, even in the event of node failures. The inherent immutability and cryptographic security of blockchain technology guarantee data integrity, preventing unauthorized alterations and ensuring transparency. Additionally, a permissioned network provides controlled access, allowing only authorized participants to join and interact with the system, thereby maintaining security and compliance. Unlike traditional databases, blockchain's consensus mechanisms and tamper-proof nature mitigate risks of data inconsistency or loss, making it a robust solution for managing critical data with enhanced reliability and scalability.

Building upon this, a permissioned blockchain network uses efficient consensus mechanisms that are not resource-intensive, differentiating it from traditional public blockchains like Bitcoin or Ethereum. Traditional blockchains often rely on proof-of-work algorithms, which require substantial computational power and energy consumption to validate transactions. In contrast, permissioned blockchains employ consensus protocols such as Practical Byzantine Fault Tolerance (PBFT) or Raft, designed for networks with known and trusted participants. These protocols enable faster transaction processing with significantly lower energy usage because they do not require complex mathematical computations to achieve consensus.

Key advantages of permissioned blockchain networks for governance in data spaces include:

- Trust and Transparency: Data governance requires clear and transparent mechanisms to manage and monitor data exchanges. Blockchain's immutable ledger provides a transparent and auditable record of all transactions, ensuring that all stakeholders have a shared, trustworthy view of the activities within the data space.

- Accountability and Traceability: For effective governance, it is crucial to know who accessed or modified data and under what conditions. Hyperledger Fabric logs every action related to contract exchanges and policies with a timestamp and the responsible entity. This traceability helps resolve disputes, enforce accountability, and ensure that all participants adhere to established governance rules.

- Compliance and Regulation: Organizations operating within data spaces must adhere to data privacy laws (e.g., GDPR, CCPA) and industry-specific regulations. Blockchain's immutable logs serve as reliable evidence of compliance, documenting how contracts and policies were created, shared, and enforced. Smart contracts can also automate compliance checks, ensuring that data usage aligns with predefined policies.

- Decentralized Oversight: In multi-organization environments, centralized control can lead to power imbalances or single points of failure. Hyperledger Fabric's permissioned blockchain model distributes oversight across all stakeholders, enabling collaborative governance while maintaining fine-grained access control.

- Data Sovereignty and Policy Enforcement: Participants in a data space often retain ownership and control over their data. Blockchain ensures that any data exchange or usage adheres to contractual agreements and policies, as these can be encoded and enforced via smart contracts. This strengthens data sovereignty while maintaining operational efficiency.

- Reduction of Governance Costs: By automating logging, validation, and policy enforcement, blockchain reduces the manual overhead and complexities typically associated with governance frameworks. The immutable logs also minimize the need for intensive audits, further lowering governance costs.

6.5.1 The DRM Blockchain approach to house cleaning.

Within the dataspaces, the Clearing House is a trusted intermediary that enhances transparency and accountability in data exchanges between participants. It securely records and verifies all data transactions,

creating an immutable audit trail that details who accessed or shared data, under what conditions, and at what time. This logging mechanism ensures compliance with data usage policies and regulatory requirements, fostering trust among data providers and consumers by providing a reliable record of all activities within the ecosystem.

Despite its crucial role, the Clearing House relies on traditional databases for storing logs and transaction records. This centralized storage approach introduces potential single points of failure, making the system vulnerable to outages and data loss. Scalability can become a challenge when handling large volumes of transactions. Additionally, traditional databases are susceptible to unauthorized alterations or cyberattacks, which could compromise data integrity and the reliability of the stored information. Therefore, adopting a permissioned blockchain network offers a superior approach, addressing these limitations by providing decentralized, tamper-proof storage and enhanced scalability. Additionally, due to the blockchain network's inherent integrity, there is no need for additional checks or validations between Clearing Houses, as the decentralized ledger provides a single, tamper-proof source of truth for all participants. Moreover, permissioned blockchain networks utilize efficient consensus mechanisms that enable rapid validation of transactions among known and trusted participants, reducing computational overhead compared to public blockchains. They offer fine-grained access control, allowing organizations to precisely define who can read or write data on the blockchain, which aligns with the varied access needs within data spaces. Furthermore, their scalable and modular architecture supports a wide range of use cases and facilitates seamless integration with existing systems, making them highly adaptable to different organizational requirements.

Additionally, current Clearing House implementations typically do not provide endpoints for integrating logs from other tools used within the dataspace framework. This limitation restricts the ability to achieve a comprehensive and unified audit trail encompassing all components of the data exchange ecosystem. In contrast, a permissioned blockchain network solution offers this functionality by supporting the logging of various tools within the DS2 framework. By enabling seamless integration of logs from multiple sources, the blockchain network ensures a unified, tamper-proof ledger that enhances transparency, accountability, and compliance across the entire dataspace. This holistic approach not only simplifies monitoring and auditing processes but also strengthens the overall security and reliability of the data exchange environment.

6.5.1.1 Assumptions/ Requirements

The DS2 DRM[18] module is intended to operate alongside the dataspace connector, with seamless integration being critical to capturing all logs and events generated within the dataspace. By closely interfacing with the connector, the DRM module ensures comprehensive logging and monitoring of all data exchanges, facilitating a unified view of activities within the dataspace. It will communicate with the policy enforcement tool to provide crucial information necessary for effectively enforcing the dataspace's policies. Additionally, the DRM module will provide an API for other DS2 tools to log activities, enabling a unified and seamless logging infrastructure across the entire dataspace framework.

In cases where a Clearing House already exists, the data stored in its common databases can be transferred to the DRM component, which offers enhanced security and reliability through its permissioned blockchain network. By adding the logs and transaction records to the DRM module, organizations can benefit from tamper-proof, immutable storage that mitigates risks associated with centralized databases, such as single points of failure and data tampering. This transition ensures that all historical data is preserved within a more

[18] DS2 D2.2 Requirements, baselines, KPI's, architecture & specifications presents the module in detail.

robust and secure framework, providing an improved audit trail, and fostering greater trust among participants in the data exchange ecosystem.

For inter-dataspace operations, the DRM offers a solution by employing an instance as a trusted intermediary that will be deployed on the DS2 trusted environment. This DRM Intermediary acts as a central mediator, facilitating the connection of DRM clients blockchain networks in a federated environment, which enables the logging and DRM of transactions across dataspaces.

Lastly, it is important to emphasize the significance of thorough validation and strict adherence to all necessary protocols when sharing data between dataspaces. Respecting data privacy laws, security standards, and agreed-upon policies is essential to maintain data integrity and trust among all participants. The DRM application will align with the Clearing House's practices by ensuring that all data exchanges comply with these requirements. This alignment promotes transparency and security, making data sharing responsible and reliable across the dataspace ecosystem.

7 CONCLUSION

The first deliverable of WP3 has systematically explored critical aspects of B2B data sharing, focusing on understanding the regulatory landscape and developing initial approaches to address complex data sharing challenges. By examining the interconnections between regulatory requirements, user perspectives, and technological solutions, this document establishes foundational insights for DS2 development.

Our initial research has established foundational frameworks for understanding data sovereignty, risk management, and trust-building mechanisms. The proposed instruments for contract management, house clearing, and policy enforcement represent preliminary steps towards creating more transparent and secure data sharing environments.

The accompanying survey marks an important methodological approach to capturing organizational perspectives on data sharing barriers from the human perspective. Its insights will directly inform future deliverables, enabling more targeted strategies for lowering resistance and fostering collaborative data ecosystems.

As the first in a series of deliverables, this document provides a critical initial mapping of challenges and potential solutions, setting the stage for more refined approaches in subsequent project phases. The ongoing work will continue to evolve our understanding of how to create trusted, interoperable data sharing mechanisms that balance organizational needs with broader innovation objectives. It will also express requirements for other modules in DS2.

ANNEX A: REFERENCES

Abbas, A.E. *et al.* (2024) 'Beyond control over data: Conceptualizing data sovereignty from a social contract perspective', *Electron Markets*, 34(20), p. 2024. Available at: https://doi.org/10.1007/s12525-024-00695-2.

Affleck, P. *et al.* (2023) 'Trusted research environments are definitely about trust', *Journal of Medical Ethics*, 49, pp. 656–657. Available at: https://doi.org/10.1136/jme-2022-108678.

Agahari, W., Ofe, H. and de Reuver, M. (2022) 'It is not (only) about privacy: How multi-party computation redefines control, trust, and risk in data sharing', *Electronic markets*, 32(3), pp. 1577–1602.

Ahokangas, P. and Myllykoski, J. (2014) 'The Practice of Creating and Transforming a Business Model', *Journal of Business Models*, 2(1). Available at: https://doi.org/10.5278/ojs.jbm.v2i1.719.

AI Act | Shaping Europe's digital future (2024). Available at: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai (Accessed: 28 November 2024).

AI Principles Overview (no date). Available at: https://oecd.ai/en/principles (Accessed: 28 November 2024).

Akram, R.N. and Ko, R.K.L. (2014) 'Digital Trust - Trusted Computing and Beyond: A Position Paper', in *2014 IEEE 13th International Conference on Trust, Security and Privacy in Computing and Communications, Beijing, China, 2014*, pp. 884–892. Available at: https://doi.org/10.1109/TrustCom.2014.116.

Alves, P., Campos, P. and Oliveira, E. (2012) 'Modeling the Trustworthiness of a Supplier Agent in a B2B Relationship', in L.M. Camarinha-Matos, L. Xu, and H. Afsarmanesh (eds) *Collaborative Networks in the Internet of Services. PRO-VE 2012. IFIP Advances in Information and Communication Technology, vol 380. , Berlin*. Heidelberg: Springer. Available at: https://doi.org/10.1007/978-3-642-32775-9_67.

American, T. (no date) 'Heritage® Dictionary of the English Language'.

Amit, R. and Zott, C. (2001) 'Value creation in E-business', *Strategic Management Journal*, 22(6–7), pp. 493–520. Available at: https://doi.org/10.1002/smj.187.

Babb, P. (2021) *The Code Pillars: Trustworthiness is about doing things differently [Blog]. UK Office for Statistics Regulation Blog*. Available at: https://osr.statisticsauthority.gov.uk/blog/the-code-pillars-trustworthiness-is-about-doing-things-differently/.

Bach, K. (1997) 'The semantics-pragmatics distinction: What it is and why it matters. Pragmatik: Implikaturen und Sprechakte, VS Verlag für Sozialwissenschaften', *Wiesbaden*, pp. 33–50.

Ball, A., 2012. Review of data management lifecycle models.

Bernal, J. (2024) 'Private sector trust in data sharing: enablers in the European Union', *Data & Policy*, 2024, p. 6. Available at: https://doi.org/10.1017/dap.2024.20.

Binmore, K. (2017) 'On the foundations of decision theory', *Homo Oeconomicus*, 34, pp. 259–273.

Bobev, T. *et al.* (2023) 'White Paper on the Definition of Data Intermediation Services'. Rochester, NY. Available at: https://doi.org/10.2139/ssrn.4589987.

Bonfiglio, F. (2021) *Gaia-X: Vision & Strategy*. Available at: https://gaia-x.eu/wp-content/uploads/2021/12/Vision-Strategy.pdf.

Bräutigam, T. *et al.* (2022) 'EU Regulation Builds a Fairer Data Economy: The opportunities of the Big Five proposals for businesses, individuals and the public sector [Working Paper]', *Sitra* [Preprint]. Available at: https://www.sitra.fi/app/uploads/2022/06/sitra-eu-regulation-builds-a-fairer-data-economy.pdf.

Burr, C. *et al.* (2024) *Trustworthy and Ethical Assurance of Digital Health and Healthcare [Report]*. The Alan Turing Institute. Available at: https://doi.org/10.5281/zenodo.10532573.

Carmichael, L., Hall, W. and Boniface, M. (2024) 'Personal data store ecosystems in health and social care', *Frontiers in Public Health*, 12. Available at: https://doi.org/10.3389/fpubh.2024.1348044.

Carter, P., Laurie, G.T. and Dixon-Woods, M. (2015) 'The social licence for research: why care.data ran into trouble', *Journal of Medical Ethics*, 41, pp. 404–409. Available at: https://doi.org/10.1136/medethics-2014-102374.

CEN-CENELEC (no date) *Trusted Data Transaction - Part 1: Concepts, terminology and mechanisms*. Available at: https://www.cencenelec.eu/media/CEN-CENELEC/News/Workshops/2024/2024-01-16%20-%20Data%20Transactions/cwa-draft-part1-0-8_clean.pdf.

Chang, Y.W. (2016) 'Influence of the principle of least effort across disciplines', *Scientometrics*, 106(3), pp. 1117–1133.

Cheshire, C. (2011) 'Online Trust, Trustworthiness, or Assurance?', *Daedalus*, 140(4), pp. 49–58. Available at: https://doi.org/10.1162/DAED_a_00114.

Cheung, A.S.Y. (2024) 'From Data Subjects to Data Sovereigns: Addressing the Limits of Data Privacy in the Digital Era', in A. Chander and H. Sun (eds) *Data Sovereignty: From the Digital Silk Road to the Return of the State (New York, 2023; online edn, Oxford Academic, 14 Dec*. Available at: https://doi.org/10.1093/oso/9780197582794.003.0005.

Commission, E. (2022) *Commission Staff Working Document on Common European Data Spaces*. Available at: https://digital-strategy.ec.europa.eu/en/library/staff-working-document-data-spaces.

Dahl, M. *et al.* (2024) *Hallucinating Law: Legal mistakes with large language models are pervasive*. Available at: https://hai.stanford.edu/news/hallucinating-law-legal-mistakes-large-language-models-are-pervasive.

Data Act | Shaping Europe's digital future (no date). Available at: https://digital-strategy.ec.europa.eu/en/policies/data-act (Accessed: 28 November 2024).

Data Act explained | Shaping Europe's digital future (no date). Available at: https://digital-strategy.ec.europa.eu/en/factpages/data-act-explained (Accessed: 28 November 2024).

Data Governance Act explained | Shaping Europe's digital future (no date). Available at: https://digital-strategy.ec.europa.eu/en/policies/data-governance-act-explained (Accessed: 28 November 2024).

Data Sharing Coalition (2021) *Data Sharing Canvas, A stepping stone towards cross-domain data sharing at scale*, *Version 1.0*. Available at: https://coe-dsc.nl/wp-content/uploads/2024/02/data-sharing-canvas.pdf.

Datos (AEPD), A.E. de P. de (2023) *Data Spaces, sovereignty and privacy by design [Blog]*. Available at: https://www.aepd.es/en/prensa-y-comunicacion/blog/data-spaces-sovereignty-and-privacy-by-design.

Davies, B.L. (2007) *Least Collaborative Effort or Least Individual Effort: Examining the Evidence*. Working Papers in Linguistics and Phonetics, No. 12; University of Leeds: Leeds, UK.

Davies, M. and contributions (no date) *from Allen, L., Sharp, M., Thwaites, E., Freeguard, G., & Baker*. Available at: https://theodi.org/insights/reports/data-assurance-white-paper/.

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) (2019). Available at: https://eur-lex.europa.eu/eli/dir/2019/1024/oj (Accessed: 28 November 2024).

Druckman, J.N. and McGrath, M.C. (2019) 'The evidence for motivated reasoning in climate change preference formation', *Nature Climate Change*, 9(2), pp. 111–119.

DSSC (2024) *Data Spaces Blueprint v1.5*. Available at: https://dssc.eu/space/bv15e/766061169/Data+Spaces+Blueprint+v1.5+-+Home (Accessed: 19 November 2024).

Duisberg, A. (2022) 'Legal Aspects of IDS: Data Sovereignty—What Does It Imply?', in B. Otto, M. ten Hompel, and S. Wrobel (eds) *Designing Data Spaces*. Cham: Springer. Available at: https://doi.org/10.1007/978-3-030-93975-5_5.

Dunning, D. and Balcetis, E. (2013) 'Wishful seeing: How preferences shape visual perception', *Current Directions in Psychological Science*, 22(1), pp. 33–37.

European Commission (2024a) 'Frequently Asked Questions - Data Act'. Available at: https://digital-strategy.ec.europa.eu/en/library/commission-publishes-frequently-asked-questions-about-data-act (Accessed: 28 October 2024).

European Commission (2024b) 'Implementing the Data Governance Act - guidance document'. Available at: https://digital-strategy.ec.europa.eu/en/library/new-practical-guide-data-governance-act (Accessed: 28 October 2024).

European Commission. Joint Research Centre. (2023) *Mapping the landscape of data intermediaries: emerging models for more inclusive data governance.* LU: Publications Office. Available at: https://data.europa.eu/doi/10.2760/261724 (Accessed: 4 October 2024).

European Data Governance Act | Shaping Europe's digital future (2024). Available at: https://digital-strategy.ec.europa.eu/en/policies/data-governance-act (Accessed: 28 November 2024).

Fassnacht, M. *et al.* (2023) 'Barriers to data sharing among private sector organizations', in *Proceedings of the 56th Hawaii International Conference on System Sciences*.

Fox, S. (2016) 'Dismantling the box: application of principles for reducing preconceptions during ideation', *International Journal of Innovation Management*, 20(6), p. 1650049.

Fox, S. (2019) 'Addressing the influence of groupthink during ideation concerned with new applications of technology in society', *Technology in Society*, 57, pp. 86–94.

Fox, S. (2024) 'Adaptive AI Alignment: Established resources for aligning machine learning with human intentions and values in changing environments', *Machine Learning and Knowledge Extraction*, 6(4), pp. 2570–2600.

Fox, S. and Rey, V.F. (2024) 'Representing human ethical requirements in hybrid machine learning models: Technical opportunities and fundamental challenges', *Machine Learning & Knowledge Extraction*, 6(1), pp. 580–592.

Frankenberger, K., Weiblen, T. and Gassmann, O. (2014) 'The antecedents of open business models: an exploratory study of incumbent firms', *R&D Management*, 44(2), pp. 173–188. Available at: https://doi.org/10.1111/radm.12040.

From the Public Sector Information (PSI) Directive to the Open Data Directive | Shaping Europe's digital future (2023). Available at: https://digital-strategy.ec.europa.eu/en/policies/psi-open-data (Accessed: 14 March 2023).

Fruhwirth, M., Pammer-Schindler, V. and Thalmann, S. (2024) 'Knowledge Leaks in Data-Driven Business Models? Exploring Different Types of Knowledge Risks and Protection Measures', *Schmalenbach J Bus Res*, 76, pp. 357–396. Available at: https://doi.org/10.1007/s41471-024-00189-z.

Glynn, I. (2010) *Elegance in Science: the Beauty of Simplicity*. Oxford: Oxford University Press.

Goodman, N.D. and Frank, M.C. (2016) 'Pragmatic interpretation as probabilistic inference', *Trends in Cognitive Science*, 20, pp. 818–829.

Graham, M. *et al.* (2023) 'Trust and the Goldacre Review: why trusted research environments are not about trust', *Journal of Medical Ethics*, 49, pp. 670–673. Available at: https://doi.org/10.1136/jme-2022-108435.

Gros, H., Thibaut, J.P. and Sander, E. (2021) 'What we count dictates how we count: A tale of two encodings', *Cognition*, 212, p. 104665.

Harvey, K. and Laurie, G. (2024) *Proxies of Trustworthiness: A Novel Framework to Support the Performance of Trust in Human Health Research*. Bioethical Inquiry. Available at: https://doi.org/10.1007/s11673-024-10335-1.

Health Data Research U. K. (hdr, U.K.). (no date) *(n.d)*. Available at: https://www.hdruk.ac.uk/wp-content/uploads/2021/09/HDRUK_TRE-One-Pager.pdf.

Heeß, P. *et al.* (2024) 'Enhancing trust in global supply chains: Conceptualizing Digital Product Passports for a low-carbon hydrogen market', *Electron Markets*, 34(10), p. 2024. Available at: https://doi.org/10.1007/s12525-024-00690-7.

Hellmeier, M. *et al.* (2023) 'Implementing Data Sovereignty: Requirements & Challenges from Practice', in *Proceedings of the 18th International Conference on Availability, Reliability and Security (ARES '23)*. New York, NY, USA, Article 143: Association for Computing Machinery, pp. 1–9. Available at: https://doi.org/10.1145/3600160.3604995.

Hummel, P. *et al.* (2018) 'Sovereignty and Data Sharing', *ITU Journal on Future and Evolving Technologies*, 1(2). Available at: https://www.itu.int/en/journal/002/Pages/11.aspx.

Hutterer, A. and Krumay, B. (2024) 'The adoption of data spaces: Drivers toward federated data sharing', in *Proceedings of the 57th Hawaii International Conference on System Sciences*. Available at: https://hdl.handle.net/10125/106926.

International Data Spaces Association (2024) *IDSA Rulebook | IDS Knowledge Base*. Available at: https://docs.internationaldataspaces.org/ids-knowledgebase/idsa-rulebook/ (Accessed: 27 November 2024).

International Data Spaces Information Model (no date). Available at: https://international-data-spaces-association.github.io/InformationModel/docs/index.html# (Accessed: 28 November 2024).

iShare Foundation (2024) *Licenses | iSHARE Trust Framework*. Available at: https://framework.ishare.eu/detailed-descriptions/functional/licenses (Accessed: 27 November 2024).

Jabeen, F. and others (2018) *Trust and Reputation Management in Healthcare Systems: Taxonomy, Requirements and Open Issues*. IEEE.

Jarke, M., Otto, B. and Ram, S. (2019) 'Data Sovereignty and Data Space Ecosystems', *Bus Inf Syst Eng*, 61, pp. 549–550. Available at: https://doi.org/10.1007/s12599-019-00614-2.

Johnston, R.G. (2004) 'Adversarial safety analysis: Borrowing the methods of security vulnerability assessments', *Journal of Safety Research*, 35(3), pp. 245–248.

Jost, J.T. *et al.* (2018) 'Political conservatism as motivated social cognition', *Psychological Bulletin*, 129(3), pp. 339–375.

Jussen, I. *et al.* (2024) 'Issues in inter-organizational data sharing: Findings from practice and research challenges', *Data & Knowledge Engineering*, 150, p. 102280. Available at: https://doi.org/10.1016/j.datak.2024.102280.

Jussen, I., Schweihoff, J. and Möller, F. (2023) 'Tensions in Inter-Organizational Data Sharing: Findings from Literature and Practice', in *2023 IEEE 25th Conference on Business Informatics (CBI)*. IEEE, pp. 1–10.

Katz, L. (2010) 'A theory of loopholes', *Journal of Legal Studies*, 39(1), pp. 1–31.

Kmenta, S. and Ishii, K. (2004) 'Scenario-based failure modes and effects analysis using expected cost', *Journal of Mechanical Design*, 126(6), pp. 1027–1035.

Koc, K. and Gurgun, A.P. (2022) 'Ambiguity factors in construction contracts entailing conflicts', *Engineering, Construction and Architectural Management*, 29(5), pp. 1946–1964.

Licato, J. and Marji, Z. (2018) 'Probing formal/informal misalignment with the loophole task', in H. Worlds (ed.) *Hybrid Worlds: Societal and Ethical Challenges, Proceedings of the*. London, UK: ; Clawar Association Ltd., pp. 39–45.

Marsh, S. *et al.* (2020) 'Thinking about Trust: People, Process, and Place', *Patterns*, 1(3), p. 2020. Available at: https://doi.org/10.1016/j.patter.2020.100039.

Martens, B. *et al.* (2020) 'Business-to-Business Data Sharing: An Economic and Legal Analysis'. Rochester, NY: Social Science Research Network. Available at: https://papers.ssrn.com/abstract=3658100 (Accessed: 28 December 2024).

Mayer, R.C., Davis, J.H. and Schoorman, F.D. (1995) 'An Integrative Model of Organizational Trust', *Acad. Manag. Rev.*, 20, pp. 709–734.

Micheli, M. *et al.* (2020) 'Emerging models of data governance in the age of datafication', *Big Data & Society*, 7(2). Available at: https://doi.org/10.1177/2053951720948087.

Mooney, J.G., Gurbaxani, V. and Kraemer, K.L. (1996) 'A process orientated framework for assessing the business value of information technology', *Advances in Information Systems*, 27(1), pp. 68–81.

Nurse, M.S. and Grant, W.J. (2020) 'I'll see it when I believe it: Motivated numeracy in perceptions of climate change risk', *Environmental Communication*, 14(2), pp. 184–201.

O'Hara, K. (2012) *A general definition of trust [Working Paper]. University of Southampton*. Available at: https://eprints.soton.ac.uk/341800/.

O'Neill, O. (2013) 'How to trust intelligently [Blog]', *TED Blog* [Preprint]. Available at: https://blog.ted.com/how-to-trust-intelligently/.

O'Neill, O. (2018) 'Linking Trust to Trustworthiness', *International Journal of Philosophical Studies*, 26(2), pp. 293–300. Available at: https://doi.org/10.1080/09672559.2018.1454637.

Opriel, S. *et al.* (2024) *Data Sovereignty in Inter-organizational Information Systems*. Bus Inf Syst Eng. Available at: https://doi.org/10.1007/s12599-024-00893-4.

Otto, B. (2022) 'The Evolution of Data Spaces', in B. Otto, M. ten Hompel, and S. Wrobel (eds) *Designing Data Spaces*. Cham: Springer. Available at: https://doi.org/10.1007/978-3-030-93975-5_1.

Perloff, R.M. (2015) 'A three-decade retrospective on the hostile media effect', *Mass Communication & Society*, 18(6), pp. 701–729.

Pettenpohl, H., Spiekermann, M. and Both, J.R. (2022) 'International Data Spaces in a Nutshell', in B. Otto, M. ten Hompel, and S. Wrobel (eds) *Designing Data Spaces*. Cham: Springer. Available at: https://doi.org/10.1007/978-3-030-93975-5_3.

Poikola, A. and Verdonck, B. (no date) 'DSSC Glossary | Version 2.0 | September 2023'. Available at: https://dssc.eu/space/Glossary/176553985/DSSC+Glossary+%7C+Version+2.0+%7C+September+2023.

//quoteinvestigator.com/2014/03/09/as-we-are/ (no date) '(accessed on 11 September 2024)'.

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

Riis, N. (2023) 'Shaping the field of EU data law', *Journal of Intellectual Property, Information Technology and Electronic Commerce Law*, 14(1), pp. 54–65.

Roschewitz, M. *et al.* (2024) 'Automatic dataset shift identification to support root cause analysis of AI performance drift'.

Rousseau, D.M. *et al.* (1998) 'Not so different after all: A cross-discipline view of trust', *Academy of management review*, 23(3), pp. 393–404.

Rulebook for a fair data economy, version 2.0 (2022) *Sitra*. Available at: https://www.sitra.fi/en/publications/rulebook-for-a-fair-data-economy/ (Accessed: 27 November 2024).

Ryan, M., Gürtler, P. and Bogucki, A. (2024) 'Will the real data sovereign please stand up? An EU policy response to sovereignty in data spaces', *International Journal of Law and Information Technology*, 32(1), p. 2024. Available at: https://doi.org/10.1093/ijlit/eaae006.

Scerri, S. *et al.* (2022) 'Common European Data Spaces: Challenges and Opportunities', in E. Curry, S. Scerri, and T. Tuikka (eds) *Data Spaces*. Cham: Springer. Available at: https://doi.org/10.1007/978-3-030-98636-0_16.

von Scherenberg, F., Hellmeier, M. and Otto, B. (2024) 'Data Sovereignty in Information Systems', *Electron Markets*, 34, p. 15. Available at: https://doi.org/10.1007/s12525-024-00693-4.

Shumailov, I. *et al.* (2024) 'AI models collapse when trained on recursively generated data', *Nature*, 631(8022), pp. 755–759.

Smart, P. *et al.* (2021) 'Risk Models of National Identity Systems: A Conceptual Model of Trust and Trustworthiness [Technical Briefing]', *The Alan Turing Institute* [Preprint]. Available at: https://www.turing.ac.uk/sites/default/files/2021-11/technical_briefing_a_conceptual_model_of_trust_and_trustworthiness.pdf.

Solan, L.M. (2004) 'Pernicious ambiguity in contracts and statutes', *Chi.-Kent L. Rev.*, 79, p. 859.

Standards, N.I. of and Technology (NIST) (2018) 'NIST Special Publication 800-37: Risk Management Framework for Information Systems and Organizations: A System Life Cycle Approach for Security and Privacy', *NIST Joint Task Force* [Preprint]. Available at: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-37r2.pdf.

Tamò-Larrieux, A. *et al.* (2024) 'Regulating for trust: Can law establish trust in artificial intelligence?', *Regulation & Governance*, 18, pp. 780–801. Available at: https://doi.org/10.1111/rego.12568.

Thornton, L., Knowles, B. and Blair, G. (2022) 'The Alchemy of Trust: The Creative Act of Designing Trustworthy Socio-Technical Systems', in *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*. New York, NY, USA: Association for Computing Machinery, pp. 1387–1398. Available at: https://doi.org/10.1145/3531146.3533196.

Verhulst, S.G. (2023) 'Operationalizing digital self-determination', *Data & Policy*, 5, p. e14. Available at: https://doi.org/10.1017/dap.2023.11.

ANNEX B: SURVEY STRUCTURE

DS2 INTER-ORGANISATIONAL DATA SHARING

Q1
Inter-organisational Data Sharing Attitudes and Practices Researcher(s): University of Southampton University email: n.s.fair@soton.ac.uk Ethics/ERGO no: 100306 What is the research about? My name is Dr Nic Fair and I am a Senior Knowledge Engineer at the University of Southampton in the United Kingdom. I am respectfully inviting you to participate in a study regarding your attitudes to inter-organisational data sharing and your organisation's current inter-organisational data sharing practices. This research is conducted as part of the EC Horizon Europe 'Data Space, Data Share 2.0 (DS2)' project which will design and develop a new modular software infrastructure to connect data sources (Data Spaces/data silos/data lakes) together for the purpose of cross-sector, inter-organisational data sharing. This study was approved by the Faculty Research Ethics Committee at the University of Southampton (Ethics/ERGO Number: 100306). What will happen to me if I take part? This study involves completing an anonymous questionnaire which should take approximately 8 minutes of your time. If you are happy to complete this survey, you will need to tick (check) the box below to show your consent. As this survey is anonymous, the research team will not be able to know whether you have participated, or what answers you provided. Why have I been asked to participate? You have been asked to take part because you have knowledge, experience, expertise and/or decision-making responsibility in the field of data sharing between businesses operating within the European Union and Associated Countries. I am aiming to recruit 50-100 participants for this study. What information will be collected? The questions in this survey are mostly multiple choice or Yes/No questions asking for information in relation to four areas. Firstly, some brief questions relating to you and your organisation. Secondly, some questions exploring your attitude to inter-organisational data sharing. Thirdly, some questions investigating your organisation's current practices in relation to inter-organisational data sharing. Finally, a few questions concerning your knowledge and understanding of data sovereignty. Please note that in order for this survey to be anonymous, you should not include in your answers any information from which you, or other people, could be identified. What are the possible benefits of taking part? If you decide to take part in this study, you will not receive any direct benefits; however, your participation will contribute to improving knowledge in this area of research and your responses will be directly applied to the development of the technological innovations in data sharing that the DS2 project will generate. Are there any risks involved? It is expected that taking part in this study will not cause you any psychological discomfort and/or distress, however, should you feel uncomfortable you can leave the survey at any time or contact n.s.fair@soton.ac.uk to request access to support services. What will happen to the information collected? All information collected for this study will be stored securely on a password protected computer and backed up on a secure server. In addition, all data will be pooled and only compiled into data summaries or summary reports. Only the researcher and their immediate team will have access to this information. No data will be passed to any 3rd Party. The information collected will be analysed and used to inform the development of the DS2 data sharing software innovations. If appropriate, the summary results may also be published in a journal and/or presented at conferences etc. The University of Southampton conducts research to the highest standards of ethics and research integrity. In accordance with our Research Data Management Policy, data will be held for 10 years after the study has finished when it will be securely destroyed. What happens if there is a problem? If you are unhappy about any aspect of this study and would like to make a formal complaint, you can contact the Head of Research Integrity and Governance, University of Southampton, on the following contact details:

Email: rgoinfo@soton.ac.uk, phone: + 44 2380 595058. Please quote the Ethics/ERGO number above. Please note that by making a complaint you might be no longer anonymous. More information on your rights as a study participant is available via this link: https://www.southampton.ac.uk/about/governance/participant-information.page Thank you for reading this information sheet and considering taking part in this research.

O I have read and understood the information above, am aged over eighteen, and agree to take part in this survey (1)

O I do not agree to take part in this survey (2)

Skip To: End of Survey If Inter-organisational Data Sharing Attitudes and Practices Researcher(s): Univeristy of Southampto... = I do not agree to take part in this survey

End of Block: Welcome and Thank You for Taking Part!

Start of Block: About You

Q2 Do you have... (Select all the relevant answers)

☐ ...knowledge of data sharing policies and practices (1)

☐ ...responsibility for implementing data sharing policies and practices in your organisation (3)

☐ ...decision-making power over data sharing policies and practices in your organisation (4)

☐ None of the above (2)

Skip To: End of Survey If Do you have... (Select all the relevant answers) = None of the above

Q3 Does your organisation currently share data (either as data provider or data consumer) with any other organisation (inter-organisational data sharing)?

O Yes (1)

O Not sure (2)

O No (3)

Skip To: End of Survey If Does your organisation currently share data (either as data provider or data consumer) with any o... = No

Skip To: End of Survey If Does your organisation currently share data (either as data provider or data consumer) with any o... = Not sure

Q4 What is your Job Role?

O Chief Technology / Information Officer (1)

O Chief Data Officer (2)

O Data Manager (3)

O Data Engineer (4)

O Data Analyst / Data Scientist (5)

O Business Intelligence Manager (6)

O Other (Please State) (7) _____

Q5 What domain does your organisation operate in?

- O Financial Services (1)
- O Healthcare / Pharmaceuticals (2)
- O Technology (3)
- O Retail (4)
- O Manufacturing (5)
- O Construction (6)
- O Transport (7)
- O Digital Services (8)
- O Agriculture (9)
- O Education / Professional Training (10)
- O Government / Civil Service (11)
- O Entertainment / Sports / Leisure (12)
- O Other (Please State) (13) _____

Q6 Which of these best describes the primary role of your organisation in data sharing?

- O Data collector and provider to other organisations (1)
- O Data analytics service provider (2)
- O Data consumer exploiting data from other businesses (3)
- O Both data provider and data consumer (4)
- O Network orchestrator (5)
- O Altruistic organisation (7)
- O Other (Please State) (6) _____

Q7 Which of these key components of a Data Sharing Agreement do you have a good understanding of: (Please select all relevant answers)

- ☐ Purpose and Scope (1)
- ☐ Data types and formats (2)
- ☐ Data owners / ownership (3)
- ☐ IP, licenses and transfer rights (5)
- ☐ Data access, limitations and use restrictions (6)
- ☐ Security and Privacy (7)
- ☐ Compliance and Governance (8)
- ☐ Roles and Responsibilities (9)
- ☐ Duration and Termination (10)
- ☐ None of the above (12)
- ☐ Other (Please State) (11) _____

End of Block: About You

Start of Block: Attitudes to Inter-organisational Data Sharing

Q8 Which of these best matches your understanding of inter-organisational data sharing?

- O The process of making consumer personal information available to multiple 3rd Parties through internet-connected platforms (1)
- O The practice of making research data available to other investigators, institutions or the public (2)
- O The on-going exchange of business-relevant data between collaborative partners (3)

Q9 Which of these best describes your overall attitude to inter-organisational data sharing?

- O I see it as MAINLY providing critical competitive advantage and business benefits (1)
- O I see it as MAINLY constituting a significant business risk (2)
- O I see it as too complex and resource intensive to transition to an inter-organisational data sharing business model even though it could benefit my organisation (3)
- O I see it as worth the time and cost to transition to an inter-organisational data sharing business model despite the potential risks (5)
- O Other (Please State) (4) _____

Q10 Would you be more willing to share data with collaborative partners within the same industry sector, or with partners in other sectors?

- O Yes (1)
- O Maybe (it would depend on many factors) (2)
- O No (3)

Q11 In your opinion, please select the 3 most important strategic challenges to inter-organisational data sharing. (You must choose exactly 3)

- [ ] Need for long-term strategic commitment (1)
- [ ] Lack of internal and external incentives (2)
- [ ] Complexities in establishing trust relationships, collaboration networks and use cases (3)
- [ ] Complexities in placing a monetary value on the data for revenue models and scalability (4)
- [ ] Potential for data misuse by recipients (5)
- [ ] Potential for competitive disadvantage (6)
- [ ] Possibility for reputational damage (within the collaboration network) (7)
- [ ] Risk of external dependencies and/or power imbalances (8)

Q12 In your opinion, please select the 3 most important operational challenges to inter-organisational data sharing. (You must choose exactly 3)

- [ ] Lack of clarity on and/or complexity of data sovereignty aspects (1)
- [ ] Cost (time, effort and financial) of establishing data sharing infrastructure (e.g. blockchain, cloud services, interoperability...etc) (2)
- [ ] Cost (time, effort and financial) of establishing data sharing processes (e.g. data quality assurance, data management, data analysis...etc) (3)
- [ ] A lack of knowledge, skills and experience within the organisation (4)
- [ ] Complexities of Data Sharing Agreements (e.g. terms of use; access rights and authorisations; security and privacy...etc) (5)
- [ ] Potential for blurred responsibilities, decision-making and data ownership (6)

Q13 In your opinion, please select the 3 most important network / partnership considerations when deciding on inter-organisational data sharing. (You must choose exactly 3)

- [ ] Reputation and reliability of partners (1)
- [ ] Clear governance and ownership structures (2)
- [ ] 'Water-tight' Data Sharing Agreements (3)
- [ ] Equality, reciprocity and fairness (equal contributions by all network partners) (4)
- [ ] Interpersonal trust (5)
- [ ] Open and effective communication channels (6)
- [ ] Use of trusted intermediaries or 3rd parties (e.g. for certifications or audits) (7)
- [ ] Shared risks and mutual value co-creation (8)

Q14 In your opinion, please select the 3 most important data-focused technical factors when deciding on inter-organisational data sharing. (You must choose exactly 3)

- ☐ Transparency of data collection, processing and use (1)
- ☐ Robust security and privacy measures (2)
- ☐ Data quality, integrity, consistency and timeliness (in- and outward data flows) (3)
- ☐ Strict data access controls (4)
- ☐ Use of Data Trusts, data pools / silos / lakes or other legal structures (5)
- ☐ Data usability (6)
- ☐ Commonly agreed specifications or technical interoperability (7)
- ☐ Clear data lifecycle and data productisation (8)

Q15 What types of information from a data sharing partner would increase your willingness to engage in inter-organisational data sharing?

- ☐ Operational procedures for data management (1)
- ☐ Technical information on the system managing the data pipeline (2)
- ☐ Risk assessment reports (any) (6)
- ☐ 3rd Party relevant certifications (3)
- ☐ 3rd Party reputational assessments (inc. customer ratings/feedback) (5)
- ☐ Company-level ethical policies / statements on business activities (4)
- ☐ Other (Please State) (8) _____

Q16 How comfortable would you be sharing data with organisations outside your partner network (e.g. state agencies, umbrella organisations, industrial hubs...etc), if doing so would lead to industry-wide benefits and/or growth?

- O Extremely uncomfortable (1)
- O Somewhat uncomfortable (2)
- O Somewhat comfortable (3)
- O Extremely comfortable (4)

Q17 How familiar are you with the relevant EU legislation and jurisdictional legal frameworks concerning data sharing? (e.g. The Data Governance Act (2023); The Data Act (2023); GDPR (2016); UK GDPR (2018)...etc)

- O I am familiar with all of them in detail (1)
- O I am familiar with the key clauses of some or all of them (2)
- O I am familiar with the broad principles of most of them (4)
- O I am not familiar with any of them (6)

Q18 In your opinion, are laws, jurisdictional frameworks and legal contracts enough to protect your interests when data sharing?

- O Defintely not enough (1)
- O Sometimes enough (6)
- O Mostly enough (3)
- O Definitely enough (4)
- O Don't know (5)

Q19 Please state the single most important thing that would persuade you to adopt (or increase) inter-organisational data sharing with a collaborative partner (max. 10 words)

_____

End of Block: Attitudes to Inter-organisational Data Sharing

Start of Block: Your organisation's current practices

Q20 Before sharing data, how often does your organisation conduct a thorough risk assessment?

O Never (1)

O Rarely (2)

O Sometimes (3)

O Often (4)

O Always (5)

O Don't know (6)

Q21 Please select the 3 data protection measures that your organisation currently prioritises the most? (You MUST select exactly 3)

- ☐ Data anonymisation (1)
- ☐ Data pseudonymisation (9)
- ☐ Data encryption (2)
- ☐ Data access controls (3)
- ☐ Use of blockchain/DLT tools (4)
- ☐ Regular cybersecurity audits and/or risk assessments (5)
- ☐ Contractual safeguards (6)
- ☐ Employee training (7)
- ☐ Other (Please State) (8) _____

Q22 To date, has your organisation established any data sharing infrastructure (e.g. blockchain, cloud services, interoperability, databases...etc)?

- O Yes (1)
- O No (2)
- O Don't know (3)

Display This Question:

If To date, has your organisation established any data sharing infrastructure (e.g. blockchain, clou... = Yes

Q23 If you are comfortable doing so, please write a short description of the data sharing infrastructure your organisation has implemented to date, otherwise just add N/A.

_____

_____

_____

_____

_____

Q24 To date, has your organisation established any data sharing processes (e.g. data quality assurance, data management, data analysis...etc)?

O Yes (1)

O No (2)

O Don't know (3)

Display This Question:

If To date, has your organisation established any data sharing processes (e.g. data quality assuranc... = Yes

Q25 If you are comfortable doing so, please write a short description of the data sharing processes your organisation has implemented to date, otherwise just add N/A.

_____

_____

_____

_____

_____

Q26 Do you (or your organisation) currently have a clear understanding of the monetary value of your data?

- O Yes (1)
- O No (2)
- O Don't know (3)

Q27 Does your partnership network currently have clear data sharing governance and ownership structures?

- O Yes (1)
- O No (2)
- O Don't know (3)
- O Not in a collaborative network (4)

Q28 Does your organisation currently have, or has it ever had, a data sharing training programme for senior management teams and the relevant staff?

- O Yes (1)
- O No (2)
- O Don't know (3)

Q29 What types of data do you share? (Please select all relevant answers)

- ☐ Company demographics (e.g. size, industry, location...etc) (1)
- ☐ Production data (e.g. materials, design, costs, downtime...etc) (2)
- ☐ Sales data (3)
- ☐ Marketing data (e.g. website hits, marketing campaigns, sector trends...etc) (4)
- ☐ Strategic data (e.g. competitive intelligence, financial exchanges, recruitment, goals and objectives...etc) (5)
- ☐ Customer data (6)
- ☐ Other (Please State) (8) _____

Q30 What formats of data do you share? (Please select all relevant answers)

- ☐ CSV (1)
- ☐ Excel spreadsheets (2)
- ☐ JSON (3)
- ☐ XML (4)
- ☐ SQL or other database specific formats (5)
- ☐ Text (6)
- ☐ Image and/or Video (7)
- ☐ PDF (8)
- ☐ Other (Please State) (9) _____

Q31 What kind of licenses do you use for the data you share? (Please select all relevant answers)

- ☐ Creative Commons (1)
- ☐ Open Data Licenses (2)
- ☐ Open Database Licenses (3)
- ☐ Thematic Licenses (e.g. machine learning or blockchain specific licenses) (4)
- ☐ Commercial Licenses (5)
- ☐ Don't know or N/A (6)
- ☐ Other (Please State) (7) _____

Q32 What kind of restrictions do you use for the data you share? (Please select all relevant answers)

- ☐ Time limits (1)
- ☐ Geo-location restrictions (2)
- ☐ Internal usage only (3)
- ☐ Access restrictions (5)
- ☐ Event or Role based restrictions (8)
- ☐ Non-aggregation / enrichment (9)
- ☐ Storage restrictions (10)
- ☐ No restrictions / limitations (6)
- ☐ Other (Please State) (7) _____

Q33 How likely are you to increase your organisation's data sharing activities in the next year?

- O Extremely unlikely (1)
- O Somewhat unlikely (2)
- O Neither likely nor unlikely (3)
- O Somewhat likely (4)
- O Extremely likely (5)

End of Block: Your organisation's current practices

Start of Block: Data Sovereignty

Q34 Which of these best describes your understanding of data sovereignty? (You may only select 1)

- O The idea that data is geolocated and therefore subject to the ethics, laws and regulations of a particular nation or jurisdiction (1)
- O A spectrum of approaches adopted by different nation states to control data generated in or passing through national internet infrastructure (2)
- O The ability of the data owner to decide how to share and use its data (3)
- O A set of core principles and actions with the main objective of establishing trust in data collection and use (4)

Q35 To what extent do you agree with this statement: *"Data sovereignty complexities significantly impact my willingness to share data."*

- O Strongly agree (1)
- O Somewhat agree (2)
- O Somewhat disagree (3)
- O Strongly disagree (4)

Q36 To what extent do you agree with this statement: *"A lack of clarity over what data sovereignty really means for my organisation significantly impacts my willingness to share data"*

- O Strongly agree (1)
- O Somewhat agree (2)
- O Somewhat disagree (3)
- O Strongly disagree (4)

End of Block: Data Sovereignty