DELIVERABLE D2.1 – DS2 PROBLEM SPACE DEFINITION

PROJECT ACRONYM:	DS2
PROJECT TITLE:	Data Space, Data Share 2.0
GA NUMBER NO.	101135967
WEBSITE:	www.dataspace2.eu
DUE DATE OF DELIVERABLE:	30.6.2024
SUBMISSION DATE:	28.6.2024
LEAD BENEFICIARY:	VTT
LEAD AUTHOR:	Jarno Halme (VTT)
REVIEWERS:	Stefano Modafferi (UOS), Soumya Kanti Datta (Digiotouch)
TYPE:	R
DISSEMINATION LEVEL:	Public

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them. Grant Agreement (GA) #101135967.

DISCLAIMER

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or granting authority. Neither the European Union nor the granting authority can be held responsible for them.

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This Deliverable contains original unpublished work except where clearly indicated otherwise.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.

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

DOCUMENT HISTORY

VERSION	DATE	DESCRIPTION	NAME	ORG
V0.1	1.2.2024	Table of contents & Structure for the document	J. Halme	VTT
V0.2	22.4.2024	Document content created based on plenary discussions and materials	J. Halme, P. Pussinen, JP. Soininen	VTT
V0.3	10.5.2024	Review of content & additions	S. Campbell, E. Salant	ICE, IBM
V0.4	24.5.2024	Content additions	J. Halme, P. Pussinen, E. Peltonen, J. Suksi	VTT
V0.4	24.5.2024	Review & Comments	S. Kanti Datta	DIGI
V0.4	24.5.2024	Review & Comments	S. Campbell	ICE
V0.5	14.6.2024	Deliverable review	S. Kanti Datta	DIGI
V0.5	14.6.2024	Deliverable review	S. Modafferi	UOS
V0.6	26.6.2024	Finalization	P. Pussinen, J. Halme	VTT
v1.0	27.6.2024	Final	J.Halme, P.Pussinen	VTT

TABLE OF CONTENTS

EXECUTIVE SUMMARY

This document details the problem space related to the inter data space interactions. The term data space has many definitions since different communities are still developing technological advancements for secure and sovereign data sharing. As data spaces are meant for a group of organizations to enable secure data sharing, there is a need to explore the next step for this advancement which is the data sharing between different data spaces. The different technological and trust-related challenges related to data sharing between data spaces are the key issue for the DS2 project.

The document is based on a desktop study examining different data spaces definitions, technologies and identifies challenges related to data spaces. The document illustrates the basic components of data spaces and expands the problem space of accessing and sharing data between the participants of different data spaces. While there are a lot of communities formed around the topic of data spaces, they are not addressing the inter data space interactions. Furthermore, the different community actions are developing all the time, so continuous monitoring of these community actions is necessary for the DS2 project.

Inter data space interaction is not yet well defined and the data space concept is under heavy development. There are many challenges or problems related to interaction between different data spaces as they can be implemented with different technologies that may be incompatible or of different technological readiness levels. Many of the roles and functions needed to enable the inter data space interactions have not yet been defined. Therefore, the DS2 project needs to align itself so that it can address at least partly these challenges and aim to contribute to the European data space development. This also presents the DS2 project a novel opportunity to contribute to the development of data spaces by providing tools for inter data space interactions and in-data space enablement. Also suitable pre-existing technological standards and community advancements are presented to be further used by the DS2 project.

1. INTRODUCTION

1.1 DS2 project introduction

The DS2 project 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.

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 (Intersector Data Space Toolkit). 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, (City Scape, Green Deal, Precision Agriculture) and case cutting across all three use cases. 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 realize cross-sectorial data driven applications.

1.1 Problem Space Deliverable Introduction

This document focuses on definitions needed to carry out the DS2 project. Definitions include common vocabulary to be used throughout the project, using existing literature and projects as starting point. Document will also describe the selected use cases to DS2 that will be used as examples when creating inter data space connectivity solutions as per targets of the project. The document is not intended to define what is needed by the use cases, nor providing solutions to recognized challenges, rather define on what the later work in this project should focus their efforts on. It can be noted though that use cases might be in different life cycle stages (especially related to data lifecycles) and are evolving also during the project. Therefore, the project must continuously reflect its own work against the use cases and not only rely on this document although the document will create a starting point for the work.

This deliverable lays down the foundations for what data space as a concept means for the DS2 project. It will also address the questions outlined in the task description related to the problem space of data spaces by:

- Detailing the data lifecycle challenges and requirements
- Assessing the data space governance and FAIR principles for DS2
- Addressing relevant specifications and standards
- Introducing partner specific background and base knowledge

The document is structured as follows: First is defined the current state of data spaces (without DS2, i.e., without need of connecting multiple data spaces). Next the single data space definitions are mapped to DS2 context, what kind of implications the known problematics of data spaces might have in inter data space context. Finally, the DS2 problem space is discussed, what services an inter data space environment requires.

2 INTRODUCTION TO DATA SPACES

2.1 Data spaces

Data spaces is an umbrella term corresponding to any ecosystem of data models, datasets, ontologies, data sharing contracts, and specialized management services (i.e., as often provided by data centres, stores, repositories, individually or within "data lakes"), together with soft competencies around it (i.e., governance, social interactions, business processes). The term "data spaces" is used widely in both academia and business. The term however lacks universal consensus and has multiple definitions. To describe a situation when a data space might be considered as potential solution or an answer to problem, the following conditions should be met:

- Data is spread amongst multiple participants/owners
- Need to establish trust in data sharing between multiple participants
- Data sovereignty and control is important

On a general level, data spaces are understood as facilitating data integration among different actors whilst supporting actors in value creation at the same time [1]. Common characteristics for data spaces according to the EU are [2]:

- Open for the participation of all organisations and individuals
- A secure and privacy-preserving infrastructure to pool, access, share, process, and use data
- A clear and practical structure for accessing and using data: common European data space have fair, transparent, proportionate, and non-discriminatory access rules, due to well-defined and trustworthy data governance mechanisms
- Respect EU rules and values, especially personal data and consumer protection, and competition law
- Enable data holders to grant access to or to share certain personal or non-personal data
- Empower data holders to make their data available for reuse for free or against compensation

In table 1, different definitions of a data space from ongoing initiatives across Europe are presented.

Source	Definition
Gaia-X	*"The term 'data space' refers to a type of data relationship between trusted partners who adhere to the same high-level standards and guidelines in relation to data storage and sharing within one or many Vertical Ecosystems."*
BDVA	*Data spaces (v1)* *Ecosystem of data models, datasets, ontologies, data sharing contracts and specialised management services (i.e., as often provided by datacentres, stores, repositories, individually or within 'data lakes'), together with soft competencies around it (i.e., governance, social interactions, business processes)* *European governed data space (v2)* *Singular but federated virtual space connecting several other interoperable spaces*

IDSA[1]	*A data space is a secure and standardized digital infrastructure that enables trusted data exchange and data-based services among various stakeholders.*
Data Space Support Centre (DSSC)	*A distributed system defined by a governance framework that enables secure and trustworthy data transactions between participants whilst supporting trust and data sovereignty. A data space is implemented by one or core infrastructures and enables one or more use cases.*
European Commission / Common European Data Space	*"Single common European data space", with a conceptual definition as "a seamless digital territory on a scale that enables the development of new data-based products and services".*

Table 1. Different definitions for data spaces

For the DS2 project, it's assumed that the term data space will be defined based on the DSSC glossary definition [4]:

"A distributed system defined by a governance framework that enables secure and trustworthy data transactions between participants whilst supporting trust and data sovereignty. A data space is implemented by one or core infrastructures and enables one or more use cases. "

This definition should fit DS2s purpose well, since the DSSC should be compatible with the existing data space initiatives described in table 1.

2.2 Elements of a Data Space

The elements of data spaces need to be explored in more detail to better understand the problem space which the DS2 project is facing. The project distinguishes a data space into different domain elements as presented in Figure 1.

[1] https://docs.internationaldataspaces.org/ids-knowledgebase/v/how-to-build-data-spaces

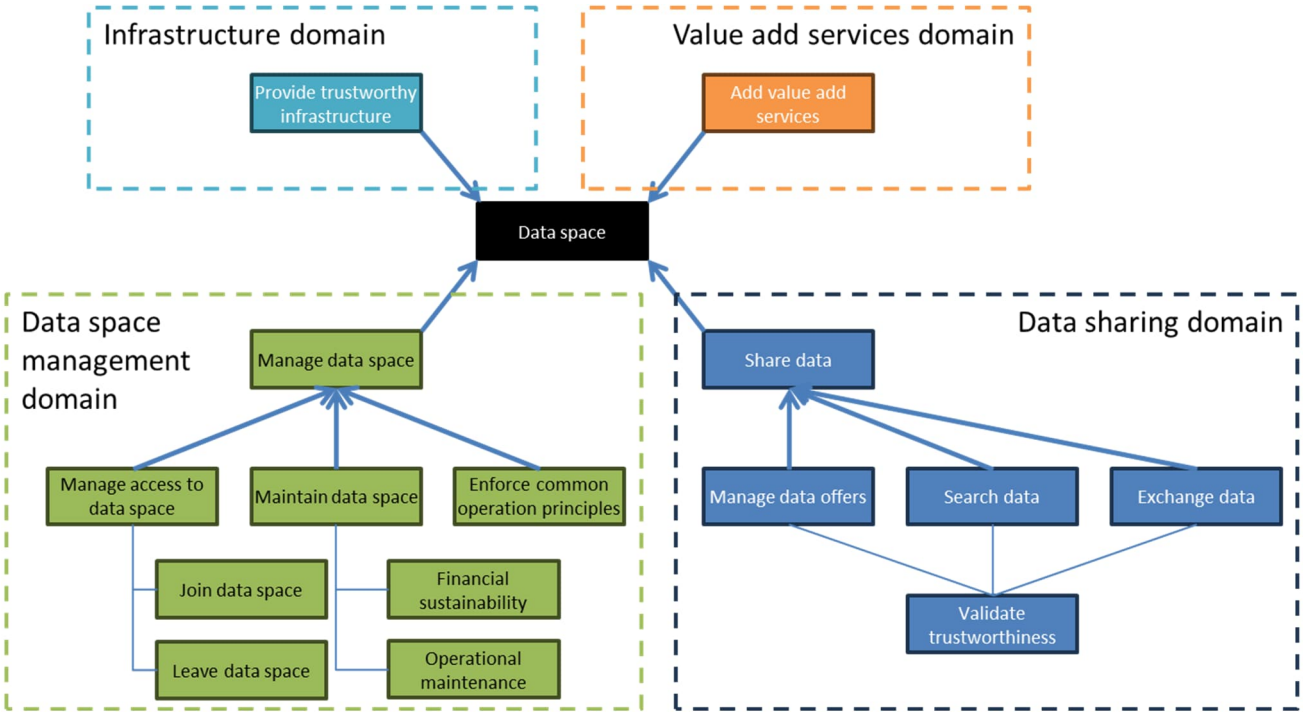

Figure 1. Elements of a data space

Based on the DSSC approach of building blocks of a data space, the different domains are:

- Data space Management domain: management and business topics related to data spaces
- Data sharing domain: technologies related to discovering data, establishing trust and exchanging the data
- Infrastructure domain: infrastructure to establish trust for data spaces
- Value added services domain: services that support the use of data space's data sharing.

The elements of a data space presented in Figure 1 above is a logical model of functionality within a data space. The services often span over multiple elements and therefore complete services may not identify to a single element. As an example, data space policies would largely be in the data space management domain, in operational management, access management (user approval) and eventually enforcement. Policies regarding data offerings locate in data sharing agreement, mainly in manage data offers -element in the data sharing domain.

For the DS2 project the data sharing domain is a critical focus point, as the core of the project is to enable connections between different data spaces. However, it cannot be overlooked that some of the technological functionalities enabled by the DS2 project may contain elements from other domains such as the management domain. As the DS2 project focuses on enabling data sharing between data spaces, functionalities may also fall into other domains as well, such as on-request data transformation functionalities would belong to "exchange data" element, while gaining access to a specific data space would deal with likely "join data space" and "validate trustworthiness" elements. Also, for the impact and exploitation of the project results, the value-added services in data spaces is a relevant domain.

The scope of functionalities provided by the DS2 project is illustrated in figure 2. In this figure, there are two different data spaces and a system belonging to each one. The aim of the DS2 project is to enable the

interaction between these two data spaces so that data can be shared between these two. To better understand the roles of DS2, the data space concept needs to be explored further.

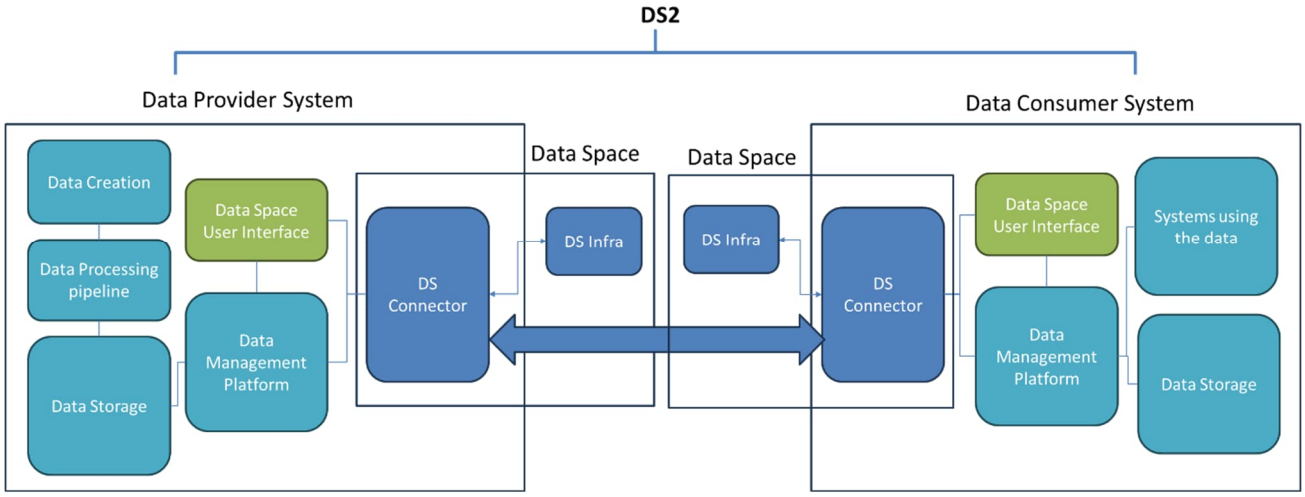

Figure 2. Data pipelines over data spaces

To acknowledge how data spaces function, the concept of data space connectors is introduced, which are also visible as the main interface towards a data space in figure 2 (the DS connector). Data space connectors are essential in creating trust and interoperability for sharing and exchanging data within a data space [5]. Instead of building individual implementations or reference implementations, the interoperability of data spaces is based on connectors with multiple levels of interoperability. As stated in the IDSA Data Connector Report [5]:

"A data connector essentially realizes two relevant aspects: It provides Data Exchange Services that are (1) the Application Programming Interface (API) to other participants in a data space to achieve interoperability and (2) the trustworthy component to handle data by implementing policy enforcement mechanisms and a common baseline for cybersecurity. However, as data can be different and the requirements for data sharing can be different as well, variants of connectors are needed."

The data space connector will cover all participant specific data sources and the connector will function as the main interface towards the data space. In figure 3, a basic structure of a data space is outlined. It shows the different logical component of an actor that in the dataspace's terminology is known either as a data provider or data consumer according to the role they play in the data exchange. The data space connector acts as the interface towards a data space, which is illustrated by the dotted line. Note that data sources are not a part of the data space, only the connector is. The connector will then provide the trust and security needed for a data space. Within the data space, the data space infrastructure represents different services needed in the data space, like the IDSA Dynamic Attribute Provisioning Service (DAPS) for connectors to connect with another connectors, broker to provide metadata and service metadata for the data space participants and data space governance service which provides the enforcement of governance principles for the said data space. The connectors manage the data consumed or provided by the owner of the connector. A data provider defines all data offerings in the own connector and the connector makes the published offerings visible in the data space's catalogue (in Broker). A data consumer browses other providers' data offerings using own connector, which downloads the data space level catalogue and can establish connections to other connectors for data download.

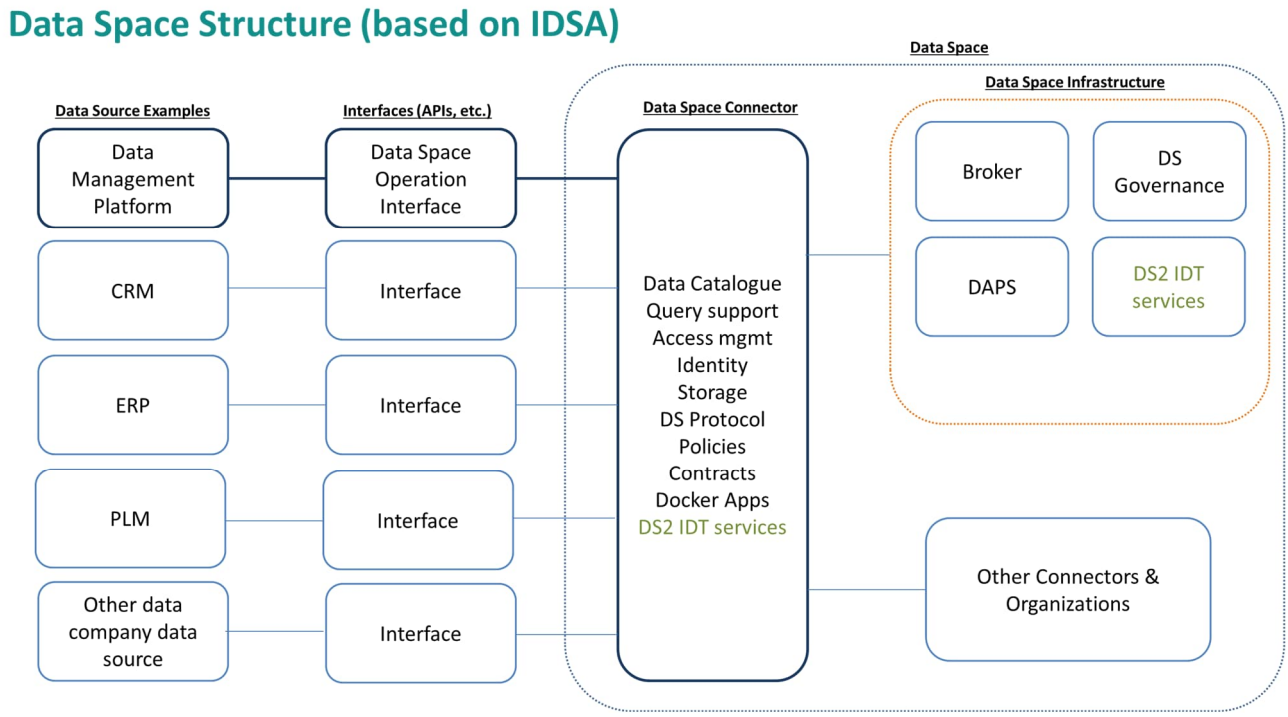

Figure 3. Structure of a data space with a data space connector

There are multiple categorizes of implementation for connectors: frameworks, open source based generic solutions, proprietary generic solutions, and integrated connectors in data related products [5]. In figure 3, the positioning of the toolkit (the DS2 IDT contains the modules defined in the DS2 project) oriented towards inter data space sharing and associated functionalities of DS2 are illustrated in green and are located within the connector and as data spaces services. DS2 also provides in data space enablement where the modules provided have a software link to the data space infrastructure and could just be used by participants to help transform or curate their existing data. How the users can access the DS2 functionalities in a data space will be detailed in the deliverable *"D2.2 Requirements, baselines, KPIs, Architecture & Specifications"*.

2.2.1 Inter Data Spaces Definition in DS2

As per the data space definition adopted above, the data spaces are typically created to support data ecosystems. Data ecosystems are typically formed to support data exchange between named stakeholders to support certain use cases that are known when setting up a data ecosystem. In case there is a new use case that requires data that is already existing and potentially shared in separate data spaces, it can be financially (and probably also technically) inefficient to form a new data ecosystem and therefore new data space to support the use case. Instead, synergies between data spaces can be achieved by creating interconnection between the data spaces.

Synergies are defined by the DSSC as gained efficiency, increased impact, or other benefits of two or more data spaces working together that are greater than if the data spaces were working separately [6]. The synergies between data spaces can be enabled by common practices, communication concepts, services and/or components, which increase data space interoperability and enable harmonised processes of using different data spaces.

2.3 Identified Technical Challenges Related to Data Spaces

A challenge for data sharing between different data spaces is that common elements to support inter data space interactions are not yet available. There are no known common tools for data space operators to initiate co-operation between data spaces like publish data catalogues visible to other data spaces. Current on-going data space actions have focused on one data space specifically and how they function technically and socially. Very little attention has been imposed on inter data space interactions and based on previous studies done on data space related challenges, it's expected that the same challenges exist also between different data spaces.

To set the scene on how the DS2 functionalities can address some of the technological challenges, a list of technological challenges (based on [7]) is presented and the impact of the DS2 project is then described.

Challenge: Sharing by design. Data lifecycle management is not designed around sharing. Most data producers do not yet consider data sharing as a possibility at the data creation stage. Existing data lifecycle management models need to improve how they incorporate all relevant processes, including preparing data for sharing and finding the correct data. The DS2 project will provide functionalities for publishing, discovering and assessing the data, as well as organizing and integrating data. The project will also provide knowledge and tools to support complex data lifecycles and support the management of rights to data over complex lifecycle.

Challenge: Digital Sovereignty. The realization of a mixed data sharing space will only materialize if data producers are guaranteed to retain their rights as the original owners, enabling them to control who can use their data, for what purpose, and under which terms and conditions. The DS2 project will provide functionalities for sovereign rights and compliance assessment. Also, the project will support practitioners for governance and relevant policy regulatory compliance.

Challenge: Decentralization. Decentralized data sharing and processing architectures. The need to guarantee that data producers control their data results in setups that forego data porting favouring decentralized data storage architectures. The DS2 project provides an IDT toolkit to enable data sharing between different stakeholders.

Challenge: Veracity: Weak verification and provenance support. Data veracity remains crucial for the sustainability of data sharing ecosystems. Data in various stages of the processing chain will need to carry traceable information about its origins and operations (i.e., metadata about its raw form, algorithms, and operations it was subjected to). DS2 envisages a blockchain-based scheme for DRM, which will provide trusted and high-level credible content protection (and to track who has access to the data) or conditional traceability of violation content service.

Challenge: Security: To enable a trusted network within which closed (proprietary, personal) data can be unlocked for exchange and sharing, issues like confidentiality and digital rights management must be addressed appropriately. The DS2 Portability functionality enables securely exchanges data from one platform/service to another (e.g. data owner to DS2 service)

The DS2 project is well aligned to alleviate the challenges of the data space domain from the technological viewpoint. There are other challenges, like business and governance challenges, related to data spaces. These are acknowledged through the DSSC building blocks. The alignment of the DS2 project according to the different data space building blocks is addressed more in chapter 3.

2.4 Data Space Technology Landscape

Data spaces are an emerging technology domain [8]. There are multiple standardization efforts ongoing, (e.g., within CEN) and technologies associated with data spaces are developing rapidly. Conversely this also means

that many of the ideas, concepts, and technologies are unknown, conflicting, or at their infancy. So the DS2 project needs to take into consideration both the potential lock-in and genericity while developing the DS2 functionalities. The technology landscape is based on the technology landscape of data spaces drafted by Sitra [7], illustrated in figure 4.

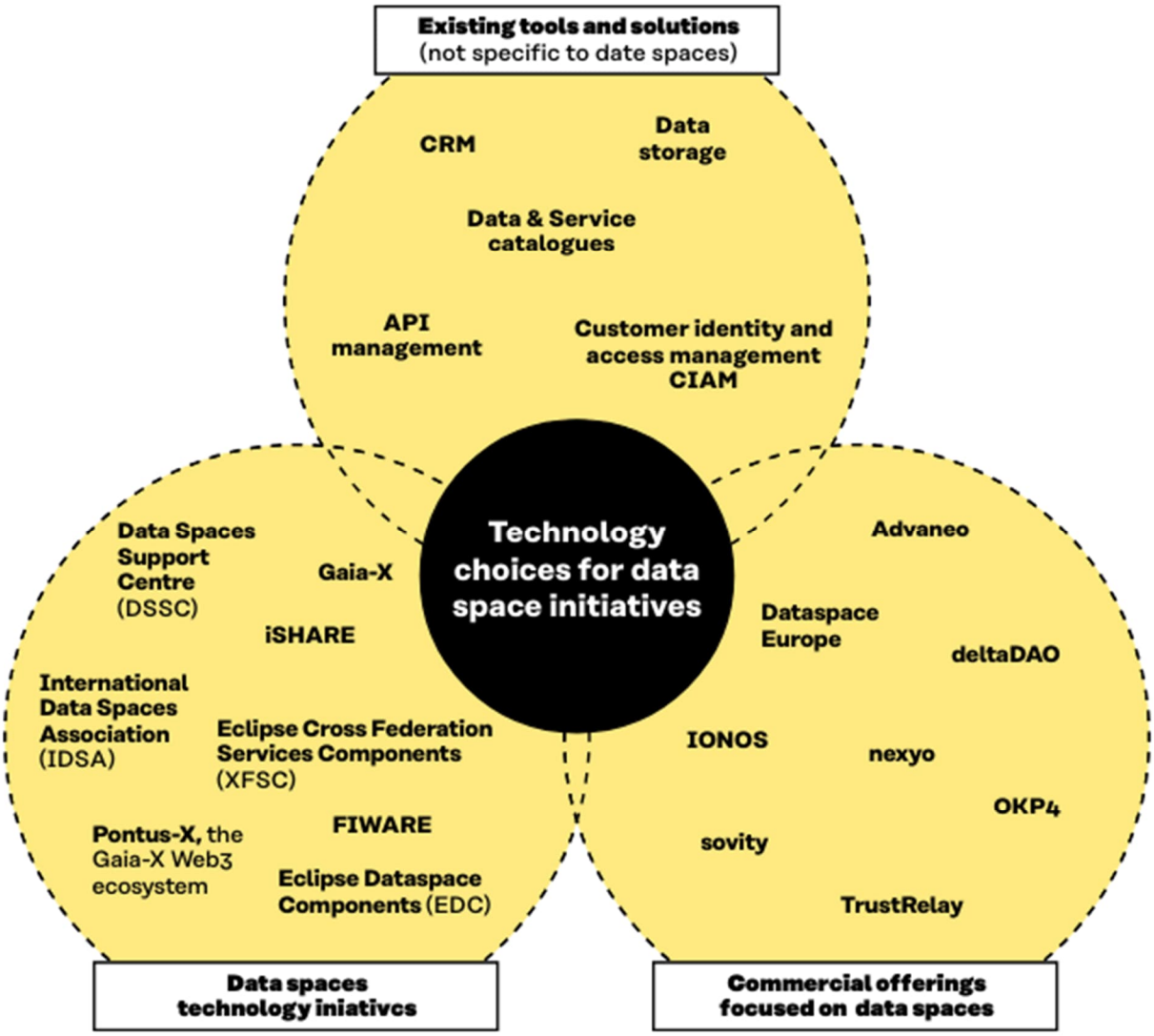

Figure 4. Technology Landscape of data spaces (adapted from [8])

In this deliverable, existing data space initiatives from the technology landscape are examined in more detail. The selected, important technology and commercial initiatives to establish a data space ecosystem are those identified in table 2 below and explored thereafter, including a description what the initiative could bring to the DS2 project.

Different Implementation Approaches	Relevance for DS2
IDSA	• Reference Architecture • Data Space Protocol (v1 and v2) • Data Space Connectors
Gaia-X	• Trust framework • Reference architecture • Component specifications
FIWARE	• FIWARE Data Space Connector • Smart Data Models
DSSC	• Glossary of terms, • Conceptual model • Blueprint • Updated list of technological standards • Data Space Building Blocks
EDC	• EDC Data Space Connector • Reference models for data space components
SIMPL	• Open source implementation of data space components (middleware)
Commercial Data Space Implementations	• Various implementation models and approaches based on proprietary assets

Table 2. Selected relevant technology landscapes for DS2 project

Data exchange plays a pivotal role in driving innovation, competitiveness, and economic growth. Within the European context, two prominent initiatives have emerged with the aim of addressing challenges related to data exchange and establishing secure and reliable digital infrastructures: IDSA and GAIA-X. IDSA is a European initiative that focuses on providing a modular infrastructure to facilitate secure and controlled data exchange across different sectors. Its approach is based on data interoperability, the creation of common structures, and data exportability, aiming to foster collaboration and innovation in the European economy. On the other hand, GAIA-X is an initiative backed by a wide coalition of companies, organizations, and European governments. Its goal is to establish a competitive and secure European data marketplace, promoting innovation and competitiveness of European businesses in the global economy. In addition to improving data exchange, GAIA-X also aims to strengthen European data sovereignty and ensure compliance with data protection regulations.

The different approaches to data spaces are introduced below and explored from the viewpoint on what technologies can used by the DS2 project and vice versa can also act as a channel for contributions developed in the DS2 project.

2.4.1 IDSA

The IDSA is a non-profit organization of over 140 member organizations committed to establishing standards for data spaces. The key features the IDSA provides for creating and maintaining data space are, which are relevant for the DS2 include:

- Reference Architecture (IDS-RAM)
- Data Space Connector
- Data Space Protocol
- The IDSA rulebook
- IDSA Glossary

The IDS-RAM outlines the key components and requirements to build a data space. It provides a common language and structure for designing and implementing data spaces, and it ensures technical and semantic interoperability between different data space implementations. This includes specifications for data connectors, data models, and security mechanisms, as well as guidelines for data governance. [9]

The Dataspace Connector is an IDS connector that is currently being maintained by Sovity. The connector was originally developed at the Fraunhofer ISST. With the help of the Dataspace Connector, existing software can easily be extended by IDS connector functionalities to integrate them into an IDS data ecosystem. Furthermore, it is possible to use the Dataspace Connector as a basis for the development of own software that is to be connected to an IDS data ecosystem.[5]

The Dataspace Protocol is used in the context of Dataspaces as described and defined in the subsequent sections with the purpose to support interoperability. In this context, the specification provides fundamental technical interoperability for Participants in Dataspaces. The Dataspace Protocol family builds on protocols located in the ISO OSI model (ISO/IEC 7498-1:1994) layers, like HTTPS. The Dataspace protocol however does not address communication between different data spaces.[10]

The IDSA Rulebook serves several purposes regarding the development and operation of data spaces. The aim is to describe clearly which rules are mandatory and which are optional guidelines. This governance framework includes functional, technical, operational, and legal dimensions. [11]

The IDSA reference architecture is already mature and is increasingly adopted in European Dataspaces. It can constitute the principles and the backbone over which DS2 can implement its intended functionalities.

2.4.2 GAIA-X

Gaia-X is a governance framework that enables data transactions between different data ecosystems [2] as illustrated in figure 5. Gaia-X enables a federated and secure data infrastructure, whereby data are shared, with users retaining control over their data access and usage. It enables the creation of links between many cloud service providers in a wider, transparent and fair ecosystem to drive the European Data economy of tomorrow. Gaia -X is a non-profit association in which its members define the Gaia-X architecture & rules. Gaia-X makes and supports others to make open source implementations of its specifications illustrated in the Figure 5 below.

[2] Purpose of a data ecosystem is to enable innovation fostering, value creation and development of new business between different stakeholders by sharing and reusing data [12]

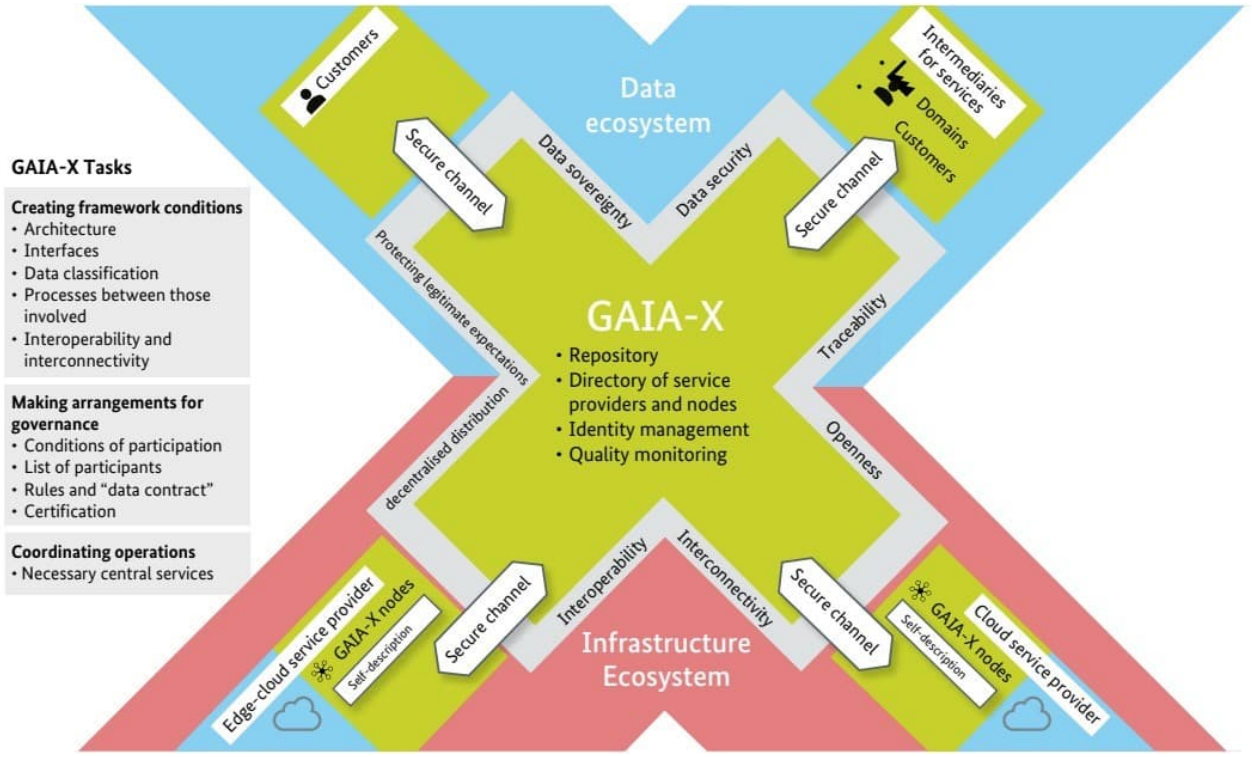

Figure 5. Gaia-X model (based on GAIA-X communication material [13])

The Gaia-X principles are:

- Federation: federated system where multiple data spaces can interact and cooperate without strict centralization. This allows each data space to maintain autonomy while participating in a larger ecosystem, facilitating interoperability and collaboration among different entities and sectors
- Decentralization: Implement distributed consensus mechanisms for decision-making within the network, ensuring no single point of control or failure. E.g., This might include the use of blockchain technologies to securely validate transactions and data in a transparent manner
- Transparency and Controllability: Ensure all operations within the data spaces are fully transparent to users and participants, allowing them to verify how their data is used and shared. Enhance controllability by enabling users to manage access to their data and control its usage
- Portability and Regulation by Automation: Facilitate data portability across different platforms and systems, ensuring that data can move freely as needed by users without losing compatibility or integrity. Implement governance policies and regulatory compliance automatically to reduce human errors and increase efficiency
- Interoperability Between Data and Services: Develop standardized interfaces and APIs to allow data and services to connect and communicate across different platforms and technologies seamlessly. This includes adopting common standards for metadata, data formats, and communication protocols
- Data Security and Governance: Integrate robust security measures to protect data against unauthorized access and attacks. Establish a clear governance framework for data management that includes privacy policies, regulatory compliance, and consent management

Adopting these Gaia-X principles into the DS2 architecture will not only provide a solid foundation for the security and efficacy of data spaces but also foster greater trust among the systems users and participants, making it more sustainable and scalable in the European context. While DS2 and GAIA-X have strong synergies to share regarding the overall goal of improving data exchange in Europe, they differ in their approach, scope,

and implementation strategies. For the DS2 project, a relevant component is the Gaia-X Trust Framework, which is the set of rules that define the minimum baseline to be part of the Gaia-X Ecosystem. Those rules provide a common governance and the basic level of interoperability across individual ecosystems while allowing users full control of their choices.[14]

2.4.3 Eclipse Dataspace Components

Eclipse Dataspace Components (EDC) started as "Eclipse Dataspace Connector" initiative, but it was soon discovered that a dataspace will need additional functionality outside the connector (although connector still being the main DS enabler) and decided to broaden the scope of the initiative to also cover these other functionalities. EDC contains components for connector, catalogue, identity hub, registration service and data dashboard. The project is open source, licensed under Apache 2.0. According to EDC internet page, also Creative Commons is an option, however Github indicates uniform use of Apache 2.0. Any organisation can take selected EDC components, complying with the license and use and develop them according to their needs. Generally, it can be noted that such a mode of operation suits organisations that 1) have the need for customisation 2) have good IT resources to perform the needed customisations or has the budget to purchase these customisations from an IT consulting company. Additionally, the organisation must be able to budget for maintenance of the environment. There are also other proprietary data space components available in the market (and more to follow). A ready solution may suite organisations that don't have too specific needs, and don't want to create and maintain customisations.

EDC is based on Gaia-X trust framework and IDSA DS protocol and therefore a data space created using EDC has good grounds on creating a data space that has potential to be interoperable with other data spaces, regardless of whether they are built with EDC or not. For the DS2 project it important to monitor the development of the different components, especially the connector and the reference models for other relevant data space components.

2.4.4 FIWARE

FIWARE Foundation is a non-profit organization that drives the definition and encourages the adoption of open standards (implemented using Open Source technologies) that ease the development of smart solutions across domains. The FIWARE Foundation, TM Forum, IUDX and OASC are leading a joint collaboration initiative to support the adoption of a reference architecture and compatible common data models that underpin a digital market of interoperable and replicable smart solutions in multiple sectors, starting with Smart Cities. Relevant outcomes from FIWARE related to DS2 include the FIWARE Data Space Connector and the Smart Data Models.

- FIWARE Data Space Connector components. The FIWARE Data Space Connector is an integrated suite of components implementing DSBA Technical Convergence recommendations, every organization participating in a data space should deploy to "connect" to a data space [15].
- Smart Data Models. The Smart Data Models initiative aims to enable actual data interoperability between diverse systems based on open-licensed data models. *"A smart data model includes four elements: The schema, or technical representation of the model defining the technical data types and structure, the specification of a written document for human readers, a URI with a working URL with basic data about the attribute or the entity, and the examples of the payloads for NGSIv2 and NGSI-LD versions."*[16]. Smart Data Models is a collaborative program to provide data models for digital twins and data spaces [17]. This can be useful for DS2 should the need for common data models in the smart city context arise, they are released as open source and are free of charge.

2.4.5 The Data Space Support Centre (DSSC)

The Data Spaces Support Centre (DSSC) is an EC project that defines a common set of rules to support data space work and creates a digital support centre for data spaces [18]. The Data Spaces Support Centre explores different needs of data space initiatives, defines common requirements and establish best practices to accelerate the formation of sovereign data spaces as a crucial element of digital transformation in all areas [19].

The DSSC provides a set of assets[3] as a starter-kit for establishing data spaces. The resources include a glossary of terms, a conceptual model for a data space, a blueprint for implementing data spaces as well an up-to-date collection of standards and technological landscape related to data spaces. For the DS2 project, these DSSC ready-made assets are important as they support most of the existing approaches to data spaces.

2.4.6 SIMPL

SIMPL is an EU project which implements open-source components for data spaces and currently cooperating with the DSSC. SIMPL provides an open source and secure middleware platform that supports data access and interoperability among European data spaces [20]. For the DS2 project, it is important to monitor to results of the SIMPL project for potential contributions from them to DS2 and vice-versa. INDRA is a member of both DS2 and the SIMPL project.

2.4.7 Other Initiatives

There are also initiatives like the Big Data Value Association (BDVA) and the Data Space Business Alliance (DSBA) that are providing support for the formulation of data ecosystems and business use of big data. These initiatives are being monitored by the DS2 project for potentially relevant channels for impact of the DS2 project results and dissemination activities.

2.4.8 Commercial Data Space Implementations

Data spaces are also implemented by commercial parties, each pursuing their own proprietary solutions. Examples include Data Space Europe and Amazon Web Services (AWS). The AWS has expanded their cloud infrastructure services to support data spaces by supporting the Eclipse Dataspace Components and FIWARE components [21].

It is always a good practice to understand what the direction of the commercial market are, and DS2 will closely monitor it. In addition, while DS2 will exploit a mixed set of the previously described open resources, it also foresees that some of the components will have a proprietary implementation. Such components might explore as an exploitation route technological partnership with other existing proprietary solutions.

2.4.9 Implementations Outside Europe

So far, focus has been mostly on European data space activities, but there are other approaches. The DS2 project will continuously monitor other data space approaches around the world for relevance; in this deliverable we illustrate other approaches from Japan and from Singapore. More approaches to data spaces are expected to emerge during course of the project.

[3] The DSSC assets are sustainable open resources that are developed and governed by the Data Spaces Support Centre. https://www.eosc.eu/sites/default/files/02_DSSC_presentation_Savvas.pdf

CADDE project in Japan

Data spaces were initiated in EU but are also being implemented elsewhere in the world. Below an example from Japan where data spaces are being implemented for a Smart City ecosystem to enhance data sharing within and between cities, this is illustrated in figure 6.

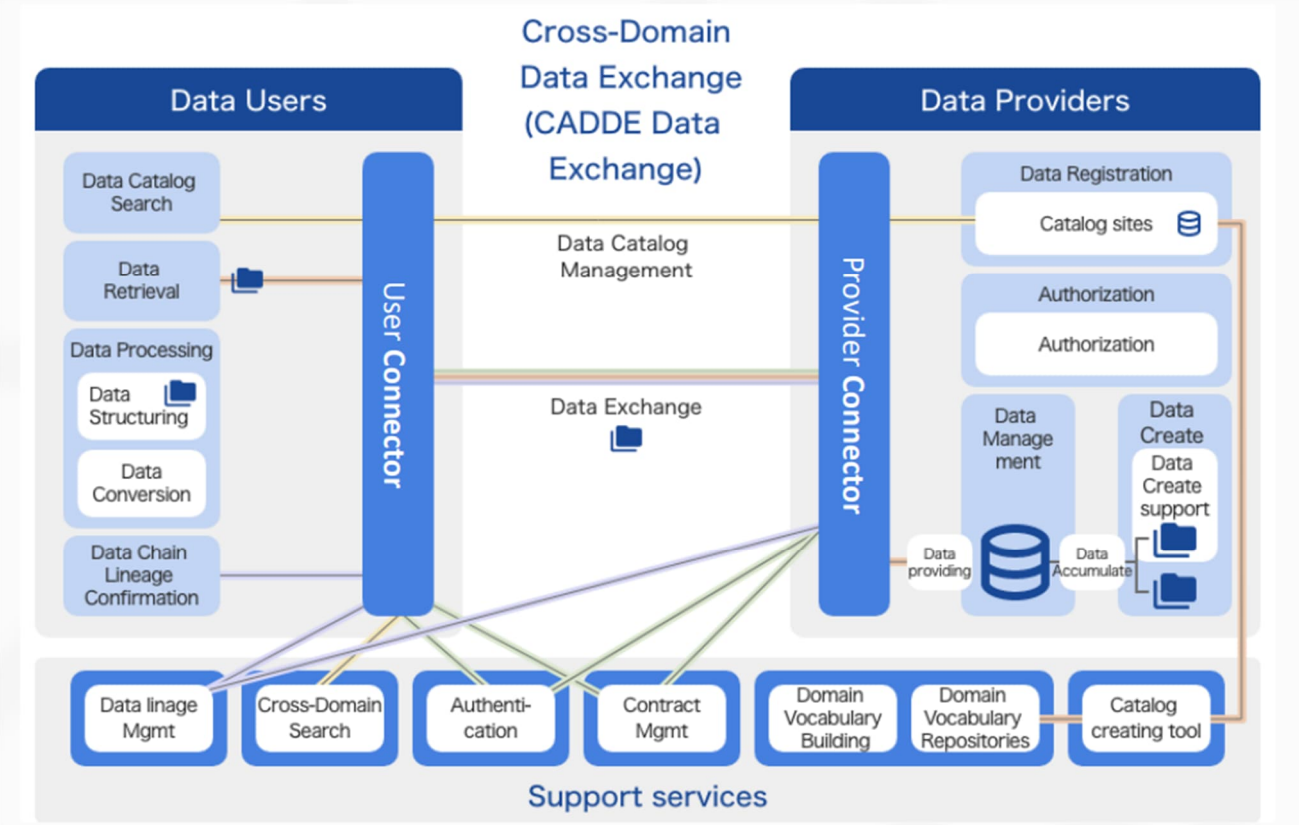

Figure 6. CADDE data space project in Japan

CADDE stands for Connector Architecture for Decentralized Data Exchange. In CADDE distributed data providers and data users participate in this network by providing connectors that serve as contact points for each other. In the data exchange between connectors, functions such as authentication and authorization, contract management, and history management are invoked and used as needed.

Trusted Data Sharing Framework in Singapore

In Singapore, the Infocomm media development authority (IMDA) has released the Trusted Data Sharing Framework to support data sharing between different stakeholders [22]. This framework enables companies to establish a set of trusted technologies and a common language for data sharing. For DS2, the envisioned data stakeholder roles in a data sharing ecosystem can utilize the IDT toolkit and modules for data preparation, management and facilitating data exchange.

The US market

At the time of writing this document, no similar initiatives were found in the US. However, the US market is home to the platform and infrastructure giants like Google, Microsoft and Amazon, who are have expressed interest towards adapting to the different European data space approaches ([21], [23], [24]).

2.5 Standards and specifications

The Data Space Support Centre has already classified relevant standards and specifications related to data spaces, based on three categories [25]:

- Data Interoperability
- Data Sovereignty and Trust
- Data Value Creation.

The above DSSC standards include in the end of May 2024 total of 124 “standards and technology landscapes” which, by being relevant in a single data space environment, are also relevant in the inter data space environment. Further, the DSSC community drives the standardization of data space. One of the first initiatives is ISO/IEC AWI 20151, which is proposed by IDSA, and backed up by DSSC work. This standard will standardize the data space protocol to be used in the data spaces.

On top of only listing potentially relevant standards DSSC also recommends certain technologies and standard to be used in named contexts. As an example, DSSC recommends using Open Digital Rights Language (ODRL) in access and usage policy creation, or Data Catalog Vocabulary (DCAT) for creating service offers.

Each of the functionalities constructed in the DS2 project may have their own specific standards and specifications to obey. DS2 module owners will investigate the relevant standards and technologies selecting the applicable ones for the modules in question (e.g., ISO 27001 for risk assessment). Additionally, there are non-data space related standards, for example for software development. Adhering to these standards is out of the scope of the DS2 project. As such, it is expected that the project participants use modern methodologies and standards to, for example, ensure quality and security of the modules.

2.6 Regulatory landscape and contractual framework

From the current regulatory landscape related to data in Europe, the following acts are currently affecting the way data is handled [26]: The AI Act, the Data Act, GDPR and the Data Governance Act. In this deliverable, the acts are shortly described, and their relevance assessed from the DS2 perspective.

2.6.1 AI Act

The AI Act is the first-ever comprehensive legal framework on AI worldwide. The aim of the new rules is to foster trustworthy AI in Europe and beyond, by ensuring that AI systems respect fundamental rights, safety, and ethical principles and by addressing risks of very powerful and impactful AI models. The DS2 project can provide functionalities that will help to achieve the AI act goals in inter data space scenarios.

2.6.2 Data Governance Act

The Data Governance Act (DGA) is a cross-sectoral instrument that aims to regulate the re-use of certain public sector data, boost data sharing through the regulation of novel data intermediaries and encourage the sharing of data for altruistic purposes. Both personal and non-personal data are in scope of the DGA, and wherever personal data is concerned, the General Data Protection Regulation (GDPR) applies. In addition to the GDPR,

inbuilt safeguards will increase trust in data sharing and re-use, a prerequisite to making more data available on the market. One of the main objectives of the DGA is to identify a new category of neutral data intermediaries that comply with a specific set of rules and are specifically recognised by the EU.

Data intermediaries under the DGA will function as neutral third parties that won't monetise the data they intermediate (e.g. by selling it to another company or using it to develop their own product based on this data) and will have to comply with strict requirements to ensure this neutrality and avoid conflicts of interest [27].

The DGA sets out specific requirements for the provision of data intermediation services. Providers of such data intermediation services will play a key role in many data spaces. A data space governance authority should also evaluate to what extent it organises any services that may qualify as data intermediation services under the DGA. For the DS2 project, this means that potential DS2 related services may fall under DGA's *data intermediation service – category* and organisations should be aware of this.

2.6.3 Data Act

Data Act complements the DGA by specifying who can create value out of data from connected products and related services [28]. It establishes clear rules for accessing and using such data within the EU, a dilemma which has been emphasized by the spreading of different sensors and related data (also known as Internet of Things). Key measures in the data act include [29]:

- Increasing legal certainty for companies and consumers engaged in data generation
- Mitigating the abuse of contractual imbalances that impede equitable data sharing
- Rules enabling public sector bodies to access and use data in exceptional cases
- New rules setting the framework for customers to effectively switch between data processing services
- A review of certain aspects of the Database Directive

As the Data Act can target data spaces, DS2 can help to comply with the Data Act regulation by providing tools and modules for data interoperability, data models and data exchange in particular. Also, the DS2 functionalities (modules) have to maintain the data act integrity and can help to maintain the data act specification via access control and interoperability.

2.6.4 GDPR

The General Data Protection Regulation is a European Union law that is applied from May 25, 2018 onwards, and requires organizations to safeguard personal data and uphold the underlying privacy rights of anyone in EU territory [30]. The regulation includes seven principles of data protection that must be implemented and eight privacy rights that must be facilitated. If you process data, you have to do so according to seven protection and accountability principles as followed:

- Lawfulness, fairness and transparency — Processing must be lawful, fair, and transparent to the data subject.
- Purpose limitation — You must process data for the legitimate purposes specified explicitly to the data subject when you collected it.
- Data minimization — You should collect and process only as much data as absolutely necessary for the purposes specified.
- Accuracy — You must keep personal data accurate and up to date.

- Storage limitation — You may only store personally identifying data for as long as necessary for the specified purpose.
- Integrity and confidentiality — Processing must be done in such a way as to ensure appropriate security, integrity, and confidentiality (e.g., by using encryption).
- Accountability — The data controller is responsible for being able to demonstrate GDPR compliance with all these principles.

The GDPR is relevant for data spaces: if a use case or transaction involves any kind of information relating to an identified or identifiable natural person (known as a "data subject' within the GDPR), the use case or data transaction participants will need to ensure compliance with data protection legislation, most notably the principles relating to the processing of personal data. This will also apply to set of mixed datasets. For the DS2 project the GDPR will impose requirements to make sure the GDPR compliance is fulfilled in data transactions between different data spaces, with modules related to identity management and traceability of data supporting the fulfilment of GDPR requirements.

2.6.5 Risks related to data spaces

There are significant risks involved related to compliance of these acts in different data spaces. As data is shared and accessed between different data spaces, risks of non-compliance with different regulatory acts increases. The DSSC has outlined the potential risks related to regulatory compliance as [31]:

- Incurrence of additional costs – the need to re-evaluate elements (related, for instance, to infrastructure, governance framework or selected technologies) brought on by insufficient consideration of regulatory requirements in a timely manner. This may lead to the incurrence of costs that could have been avoided and may temporarily disrupt the proper functioning of the data space.
- Fines and penalties as a cost – the incurrence of fines by the data space (members) as a result of non-compliance with mandatory legal requirements.
- Legal uncertainty on liability – among data space users on how to deal with specific legal issues or subpar protection of third party rights, brought on by the lack of tools, guidance or policies at the data space level.
- Invalidity of contractual framework – failure to comply with the mandatory requirement imposed by the regulatory framework or the incorrect assessment of legal relationship between the parties to a contract may result in the contract being invalid or the contract inaccurately reflecting obligations and responsibilities of the parties (e.g., some exceptions under copyright law are mandatory and cannot be overwritten by contract).

DS2 should therefore put in place risk assessment procedures that should also consider the distributed and peer to peer nature of the relationships where a trade-off between sharing information on the different systems and their business sensitivity be found.

2.6.6 Contractual framework

Data spaces are stated to have agreements on two different levels: data space agreements and data transaction agreements. The data space agreements define the commitments of the data space participants: they constitute the data space, implement the governance framework and regulatory compliance, as well as add common elements to the data space such as standard clauses and licenses. The data transaction

agreements state the data product contracts and reflect the terms and conditions on how each data product is made available. Data products bundle a set of resources (e.g., data sets and data services) into a consumable form to provide meaningful business value.

For the DS2 project, it is relevant to address the compatibility of these contracts in the inter data space environment. Partly these contracts are set out on a data space level, but the compatibility of these contracts needs to be addressed.

2.7 Existing data spaces

While more than 100 different data spaces exist for various industrial sectors (as noted in the IDSA Radar [32]), this deliverable lists the following which are relevant to the DS2 use cases and the DS2 project as a whole.

- Smart Connected Supplier Network – SCSN is a data space for sharing data in supply chains of manufacturing industry. It has several hundred members and consist of a message standard and a data space detailing how information can be shared in a secure and controlled way.
- Agdatahub – currently at implementation stage, this data space is for agricultural/agrifood sector.
- City dataspace – It utilises the reference architecture put forward by IDSA for the smart cities sector. It focuses on the challenge of rapidly increasing and enabling the usability of smart city related open and urban data. The proof-of-concept exploits mobility and geodata across several city municipalities in Europe and is currently in implementation stage.
- Datahub.tirol – it implements a data ecosystem to boost the data economy of the Tyrol region and brings together several industrial sectors (relevant to DS2 use cases) like agrifood, energy, green deal, smart cities, and more. Currently at implementation stage, this data space utilises IDSA reference architecture and the Eclipse dataspace connector.
- DIVINE – it is a dataspace stemming from the EU-funded DIVINE project creating an agri-food data space and validating that in four European pilots (Slovenia, Ireland, Greece, and Spain). It is currently transitioning from implementation to operational stage and deploys Eclipse dataspace connector to the four pilots.
- DjustConnect – The objective of this data space is making data sharing safe and efficient for farmers and horticulturists. It offers a user-friendly dashboard with guaranteed implementation of the European code of conduct for data sharing in agriculture sector.
- European data space for smart communities – It is creating a cross-sectorial (e.g., agriculture, green deal, public sector, mobility etc.) data space for Governments with an aim to deliver the best possible e-services to citizens. It considers interoperability as a part of the service delivery and achieving goals like Green Deal. It uses the reference architecture of DBSA.
- Green Deal dataspace – Positioned as an open source data sharing ecosystem, it deals with resilience and sustainable economy targeting the Green Deal/Circular Economy sector. Business end users benefit from a decentralised infrastructure operating in a protected digital space increasing business/service delivery resilience.
- ZeroW dataspace – It addresses the need to move from today's fragmented solutions targeted on isolated agri food loss and waste (FLW) problems to synergised solutions based on shared data, knowledge, and collective intelligence – breaking siloed solutions and 'opening' closed ecosystems.
- Mobility data space – It is an open ecosystem that exploits IDS-based data space and linking existing data platforms for decision support in intelligent traffic and mobility systems. This is already an operational data space, close to transitioning to a scale up version.

2.8 FAIR principles in DS2 context

FAIR stands for Findability, Accessibility, Interoperability and Reusability. Following FAIR principles is agreed in the grant agreement, while it is understood that in projects (and especially in data projects) it is exactly the principles that can be followed, not necessarily all recommendations targeted to a single organisation. There are several self-assessment tools for organisations available such as from Zenodo [33].

In the DS2 operational context, the data shared between data spaces must also support FAIR principles. DS2 functionalities should target to make the data:

- Findable: This invariably involves both finding a data space containing relevant information and then determining the datasets or finding the right data sets which in turn can be tracked back to a specific dataspace. These can be treated similarly from the findability point of view (information about a data space is a dataset). Findability is achieved by, for example, using persistent identifiers for data, creating rich metadata, using standards, inserting search keywords, using semantic search engines (or allowing existing search engines to use the metadata) and creating indexable metadata.
- Accessible: Accessibility means that after finding a relevant dataset the user can also access the actual data based on Provider and Consumer governance principles. To make data accessible the provider must consider topics such as data storage and backup, refinement rate, identity, and access management (or open data), IPRs, embargos, access protocols, format, use restrictions and rules, metadata licensing, storage duration and documentation.
- Interoperable: After the data has been accessed the user must be able to use the data. The following actions are examples of factors that can maximise interoperability: Standardised metadata models, standard formats of data (e.g., file formats), common ontologies and vocabularies (if not possible, considering translation services to other -standard- ontologies), references and provenance of the data.
- Reusable: Even if the data can be accessed and is in readable format, it might not be usable. Reusability increases by ensuring some of the attributes in the previous steps are in place, such as metadata, ontologies, vocabulary etc. Reusability can be further increased by appropriate documentation, description of refinement process steps, provenance standards, description of the quality assurance process, documenting cross-validation processes with other data sets as well as description of the sensitiveness and risks related to both data itself and regulation.

2.9 Governance requirements

This section focuses on the inter data space governance challenges. As data governance can be defined as the correct management and maintenance of data assets and related aspects, which includes data rights, data privacy and data security, the purpose of inter data space governance should be to define common principles, rules and requirements to organize access and management of data by participants according to the different roles the participants have in the ecosystem.

The first inter data space governance challenge is participation in the interconnected data spaces. The inter data space governance should provide clarity on how to participate interconnected data spaces, who can participate the interconnected data spaces and what it entails to participate the interconnected data spaces. Therefore, the issues that inter data space governance need to address are among others whether participation in interconnected data spaces require membership in the interconnected data spaces, whether participation in the interconnected data spaces require something else or whether participation in the interconnected data spaces can be allowed without any requirements. In this context, it is unclear to what

extent inter data space governance needs to manage interconnected data space stakeholders and hence stakeholder identification, definition of stakeholder roles and the relationship between different roles.

The second inter data space governance challenge is data sharing in the interconnected data spaces. Inter data space governance should determine what data can be shared to stakeholders in other data spaces and under what conditions. Therefore, the inter data space governance should take into account the governance of at least two different data spaces, interpret these governances and facilitate data sharing between data spaces under these conditions. In this context, it is unclear to what extent inter data space governance should guide individual data space governance in terms of how to enable data sharing between different data spaces, considering the rights and interests of data space stakeholder members.

The third inter data space governance challenge is protection and assurance of the rights and interest of data owners in the interconnected data spaces. In this context, it is unclear to what extent the data should be visible to the different stakeholders in interconnected data spaces, considering the rights and interests of the data owners. Thus, for inter data space governance issues such as whether the data of one data space is presented to other data spaces in a catalogue manner, such as within of one's own data space, or whether participants of other data spaces can only find out about the data of another data space by specifically requesting it would need to be addressed.

It is worthwhile for inter data space governance to address the challenges regarding participation, sharing of data and protection and assurance of the rights and interests of data owners. As the purpose of governance is ensure the protection and assurance of rights and interest of its stakeholders and to promote value creation and enable solutions that support it, it can be argued that inter data space governance would be worth having for each data space separately, with its own conditions according to the wishes and needs of its own stakeholders. This can be counted among the inter data space governance challenges that need to be addressed.

3 DS2 AND COMMON ELEMENTS OF DATA SPACES

As described in section 2, data spaces are feature-rich technical constructs within a trust framework that support data ecosystems with fair and trusted approaches to share data. As per this definition DSSC has split the characteristics of a data space to building blocks, including on one side the technical building blocks (e.g., a data space must have a modern technical framework for smooth sharing of data) as in figure 8 and on other side the organizational and business building blocks (e.g. even the best technical solution is not useful if there are no reasons or rules to use it) as illustrated in figure 7. In this section these building blocks defining a data space are discussed and their relevance to DS2 project assessed initially. All the building block descriptions are adapted from the DSSC blueprint [34] using also further refined definitions from DS4Skills Blueprint [35].

The building blocks are defined by DSSC for the establishment and management of a single dataspace. While DS2 considers intra-dataspace scenario, its objectives focus on inter data space activities. This is not an issue because principles and definitions that apply to a single data space, are largely valid in inter data spaces scenarios and DS2 will fine-tune them as appropriate.

Figure 7. The Organizational and Business Building Blocks of Data Spaces.

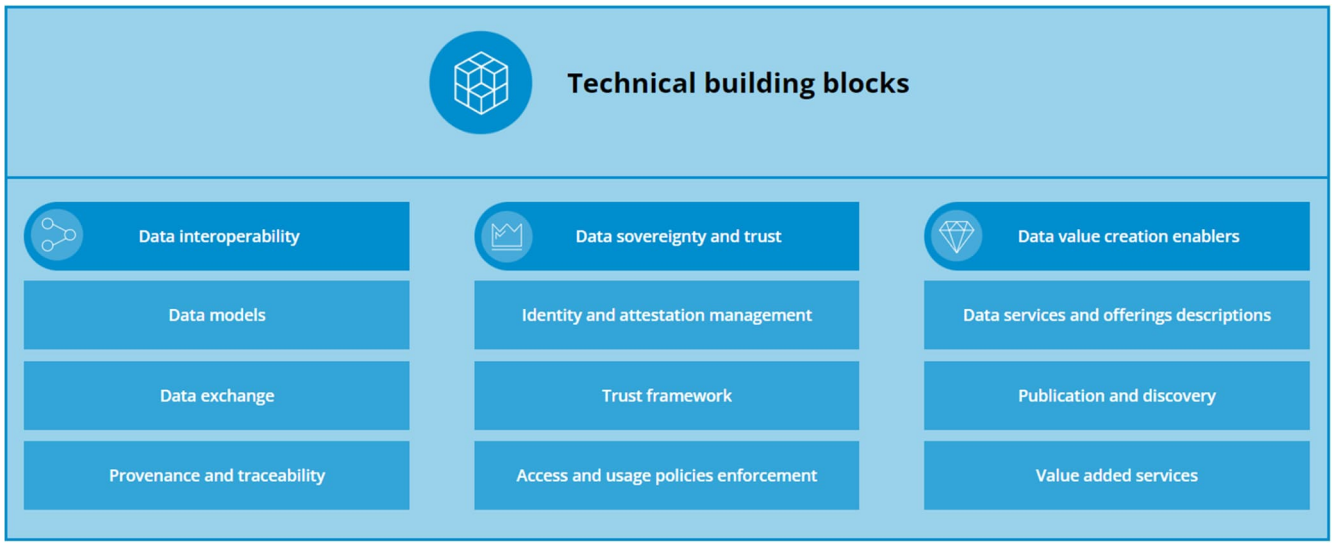

Figure 8. The Technical Building Blocks of a Data Space

Such elements can also be described as a protocol stack, where the Business, Governance and Legal (see figure 7 and 8 above) would locate on a "trust plane", the entities required to actualize the data transfer (such as discovery, transaction contract negotiation, and initiating the data transfer process) are on "control plane" and the actual data transfer often using existing data transfer methods (such as REST API) locate on the "data plane". DS2 work packages need to understand in which plane the work package is working on to target the right audience for the deliverables. In this chapter, these general data space building blocks are explored from the DS2 perspective to understand what kind of functionalities the DS2 project can bring into data spaces.

3.1 Business

This part includes all elements required to develop valued use cases within a data space. The value can be financial, social, environmental etc.

3.1.1 Business model development

In business model development it is essential to identify participants for the use case and/or data space. There must be a value proposition to each of the participants. Within a data space both revenue and cost models must be defined, and the operations of the data space agreed. These will be reflected in the contractual framework chapter (data space contract) later.

DS2 view: Investigate data-driven business models is relevant, but not core for the DS2. For this reason, the business cases defined in the use cases will only be at draft level.

3.1.2 Use case development

All data spaces are serving a purpose, a reason why they have been created in the first place. This purpose is the value set in the business model development and the way to realize this value is through use cases. Use cases define how the stakeholders use the data to create value. In the use case development phase the use cases must be identified and described. The synergies between stakeholders as well as identifying the data sources within the data space must be defined. The principles of value sharing need also be agreed by dataspace participants together with mechanisms how the adherence to these principles is ensured.

DS2 view: Despite the project targeting a relatively low TRL, the presence of Use Cases to validate the DS2 approach is key. Their definition should include and highlight the datasets used, any additional service required for their implementation, what technical challenges they are facing and how the identified solutions can scale up and be generalised. Detailed use case definitions are addressed in the deliverable D2.2.

3.1.3 Data product development

According to DSSC blueprint this is the definition of data products: "The providers of data products rely on the functional data space to deliver the data products to the users (other data space participants). These products can be: i) Raw data; ii) Products and services built using the data; iii) analysis or insights" [4]. Data products bundle a set of resources to consumable form, including technical and commercial attributes.

DS2 view: DS2 may offer services that help the users to create data products (data assets) beyond the sharing of raw data and some of the foreseen modules cover this part, also in inter data space scenarios. Creating "consumable form" between data spaces may create additional requirements which will be addressed by use cases in the DS2 project. The challenge is both technical, business, and operational and DS2 modules will focus more on the technological challenge to be faced to create data products.

3.1.4 Data space intermediary

Intermediaries can be understood as a library of data space enabling modules services that eventually form and enhance the features of a data space. The intermediaries lower the bar of creating a data space by allowing easy adoption of the required functionality to enable a data space. As a business function the intermediaries definition includes listing the required services within a data space, their roles in the business model as well as the rights and rules for each of the intermediaries.

DS2 view: DS2 may exploit the concept of intermediary to implement intra- and, above all, inter-dataspace functionalities where two distinct actors belonging to two distinct dataspaces need to exchange data products. The intermediary functions can be located within individual connectors or in data space's generic infrastructure, or even outside a data space in a federated environment. The intermediary functions can be located within individual connectors or in data space's generic infrastructure, or even outside a data space in a federated environment. According to the DS2 vision, each of the intermediary application will be independent, but the configuration and roles are defined during the DS2 project for creating a multi-purpose data space intermediary actor as one of the potential exploitation routes for the project.

3.2 Governance

Data space governance define the ways of working within a data space, including especially the organisational governance and data sharing governance. That is, not only governing the technical ways of data sharing transactions, but also organisationally for example rules on accepting new participants to the data space.

3.2.1 Organizational governance

Organizational governance sets the highest level of agreements for a data space. It defines the coordination, decision-making and goal setting within a data space to ensure all participants are treated equally. The key elements of this building block are organizational form and strategies, roles and responsibilities, governance

[4] https://www.skillsdataspace.eu/blueprint/business-building-blocks/

mechanisms and processes. The governance can also be separated to establishment and operational phases, which may often overlap in time, meaning that the rules of the data space can (and will) be iterated during operations if the preset governance rules are being followed, i.e., rules are refined in the agreed order.

DS2 view: Organisational governance is possibly one of the most challenging topics in inter data space connectivity. It should be studied if there can be a governance body for each inter data space connection, or should the trust be created with some other method to ensure free data flow between data spaces. The governance establishment should include topics as governance authority, roles, responsibilities, and rulebook creation. In inter data space environment, it should be studied if these can be standardized at least for a typical data sharing between data spaces, lowering the threshold of trouble when business needs a piece of data from another data space, after all, that is the most important reason triggering data space development in the first place. Further, during operations, governance can include dispute resolution, feedback mechanisms, rulebook maintenance, scaling, compliance enforcement, risk management, stakeholder engagement, transparency and reporting, external audits and reviews, contingency planning, following legal and regulatory compliance. In DS2 context any or all of the above can become a relevant topic during the lifecycle of the inter data space connectivity and it is recommended address the emerging topics in DS2 to ensure such problems are at least documented for further study in case they are not in the scope of DS2 project.

3.2.2 Data sharing governance

It is essential to define the ways of working on operational data sharing whereas data sharing governance focuses on the agreements of the technical data sharing operations. Defining these governance principles is a trade-off between simple implementation (low cost of both joining a data space and maintaining presence in a data space) and features within a data space.

DS2 view: Data sharing governance has utmost importance in the DS2 project. Governance agreements include take a stance on services such as identity, catalogue meta data, connector, observability, marketplace, and escrow. Some or even all these services might turn out to be DS2 modules. It is advised that the different viewpoints are studied in DS2 to gain understanding of the required services as well as addressing the targets to the DS2 project, being also able to point topics that DS2 might not be able to answer, if any.

3.3 Legal

Every participant in data space functioning over company / organisation borders must comply with possible regulations and is advised to create legally binding contracts at least when joining the data space. Contracts are created to ensure all participants are aware of the rules and responsibilities in a data space.

3.3.1 Regulatory compliance

The regulation landscape evolves constantly, and data space participants should be aware of existing and planned regulations in the environment where the data space operates. The relevant legal frameworks should be listed (and applied) when establishing the data space and the governance function of the data space should also be able to manage suggestion to changes in the regulatory environment and a participant's compliance to it. Regulatory compliance includes pragmatic categorisation of the data types and processing purposes. This is operational action that ensures compliancy with respective regulation for data handled in the data space.

DS2 view: The regulatory frameworks must be addressed in the DS2 project. It should start from understanding the relevant regulation. Given the complexity of the regulations, the approach to their assessment should be risk based rather than comprehensive. The risk-based approach provides a pragmatic way to assess the risk

connected to the potential violation of a regulation, so becoming useful as practical instrument to be performed by the data actors.

3.3.2 Contractual framework

This part includes bilateral and multi-party agreements between the participants of a data space. Contracts can define for example membership and accession agreements, the governance authority, service-level agreements, and other rules how the data space can function efficiently.

DS2 view: Smart contracting and contract enforcement are key in the current digital world. While contracts already exist in single dataspace, DS2 should understand how to extend them in inter-dataspace scenarios. Conflict management should also be considered as a potential need if different dataspaces have incompatible contractual frameworks.

3.4 Data Interoperability

This section defines the mechanisms required for data to be interoperable so that data can be consistently interpreted by different parties and thus minimising uncertainty.

3.4.1 Data models

Data understanding must be shared by data providers and consumers. Ontologies, data models, schemas, mappings, API specifications to annotate and describe the data (metadata) can be created to facilitate this understanding which can go as far as creating common vocabularies. Vocabularies are also subject to change, and data spaces must have the mechanisms to curate (add, modify, delete) them. The technical component to perform these tasks is called a Vocabulary hub. Various standards exist and must be considered (e.g., RDFS, OWL, JSON-, and SKOS).

DS2 view: The vocabularies can (and most often are) domain and industry specific. Even the same terms might have different meanings in different industries. DS2 should provide or adopt mechanisms for data space data transfers where different vocabularies are involved which could include the presence additional entities/roles to comply with the challenges when data is transferred outside its original data space. Aligning vocabularies across dataspaces by replacing one with another is very likely difficult if at all feasible for operational and business considerations. Wrappers, intermediary common ontologies, and on-the fly ontology translations/transformation are more likely to be feasible solutions.

3.4.2 Data exchange

Data exchange is defined in DSSC blueprint as “the building block providing guidelines and technical specifications about common communication protocols that all data space roles can use to transmit standardised data.”[5]. The defined capabilities include APIs, specifications for adoption of APIs as well as the tooling to maintain the APIs and is tightly linked with the data models defined in the previous chapter.

DS2 view: As the data exchange manages the data transfer intra data space, exactly the same protocols are likely to be used in inter data space connections, improving manageability of the protocols and transactions, as well as cost of implementing inter data space connections. The DS2 should use these existing intra data space standards for inter data space exchange.

[5] https://www.skillsdataspace.eu/blueprint/dssc-technical-building-blocks/

3.4.3 Provenance & traceability

Provenance and traceability are defined in DSSC as "description of how the data sharing process is being monitored within a data space."[6]. Especially in data spaces with highly regulated data, it is necessary to make the data sharing process observable. The focus is on both backward- and forward-looking evidence of origin of data and what was done with the data. Both these aspects might be regulated (depending on industry), and potentially required by both the data providers and consumers. Provenance is important for the clients to ensure the legitimate origin of the data, as well as the treatment of the data processing in the value chain. Traceability supports the assessment of what the data has been used for and where copies of the data are stored.

DS2 view: DS2 will need to extend the concept of provenance and traceability across data spaces. No solutions that would prevent provenance and traceability functions should be approved. The DS2 modules might need the architecture to support specifications mechanisms to log certain actions.

3.5 Data Sovereignty and Trust

A fundamental factor for a data space to function is the trust between data space participants and even more so between participants sharing data across dataspace. This incorporates both trusting the other participants and maintaining sovereignty over the data shared in the data space. Participants can define the policies for access and usage, though some are defined on data space level. Participants also can verify the identity of entities (people or organisations) fetching their data, and finally participants can verify that the others adhere to the set rules. Another element to ensure trust relies on a risk assessment approach which will require a certain degree of data sharing and in turn will provide information on the counterpart of the data exchange sufficient to improve the mutual trust to a suitable level.

3.5.1 Access & usage policies and control

Access & usage policies and control is essential to achieve data sovereignty. Providers may create own policies for their data, for each consumer separately. Such policies can include, for example, a limited number of times the data can be accessed (e.g., "only once") or limiting the time period a consumer can access the data (e.g., "until [date]). Such limitations can be set by the provider and/or by the data space, and they may create new policies as per their needs. The policies can be hierarchically separated to membership policies (data space level), access policies (data catalogue level), contract policies (data contract level) and usage policies (data asset level). The access and usage policies define especially the catalogue and asset level policies. That is, who can see the data in catalogue and if it is chosen to use the data, what are the conditions and restrictions.

DS2 view: When a data space is established for a certain purpose the participants define the common policies under which the data can be used, and individual participants may subset or adapt dependent on the governance of the dataspace. Policies should be expressed in machine readable format and the system should include policy enforcement modules able to enforce them. DS2 is focused on inter dataspaces scenarios and thus emphasis should be put in addressing situation where the policies of 2 different data spaces present potential conflicts that need to be resolved. Policy matching and negotiation mechanisms between data spaces should be considered as a possible solution.

[6] https://www.skillsdataspace.eu/blueprint/dssc-technical-building-blocks/

3.5.2 Identity management

Identity management includes capabilities to register, maintain and use the identity information to act in a data space in a trusted manner. Such registration might be needed for companies or organisations, people, hardware, software, data products, licenses etc. It can be noted that often humans are using discovery services to find the right data but eventually the user of the data (and the one collecting the data) is often a piece of software. Both need their own identities. Technical capabilities include the mechanisms to allow issuing, holding, and verifying identities, as well as deleting them.

DS2 view: Similar to the access and usage policies, an established data space has defined their identity management principles and methods. If standards (e.g., Oauth 2.0) are supported, mechanisms for created federation of authentication authority already exists. However, scenarios less aligned might also arise. The current message flows and standardisation (e.g., IDSA connector certificates) must be carefully analysed to ensure interoperability between data spaces without a need to create a manual integration to each of them. One may be mechanisms recognition of identity providers between DS2 enabled dataspace since theoretically then the individual participants identities are also proven.

3.5.3 Trust

Trust is defined in DSSC as "the need to provide assurance to all participants of a data space to ensure the secure and reliable exchange of data between different entities within a data ecosystem while fostering data sovereignty". It can be achieved through the implementation of a technology stack that validates for instance claims, identification, authorisation and intended use of the data, but it can also more efficiently achieved by performing risk assessment where an analysis of the risks, rather than the certification of actual facts, often dynamic and difficult to capture, achieve the desired level of mutual trust. The risk assessment approach is also better suited to capture human trust where objective and subjective element coexist and mutually influence one each other.

DS2 view: Trust is key in data sharing and DS2 needs to explore the best framework to manage different user trusts in an inter data space environment.

3.6 Data Value Creation Enablers

This section discusses the technical enablers for value creation, including awareness of the data as well as guidelines to access the data.

3.6.1 Data, services, and offerings descriptions

The data offering is described in a standardised way using best practices (such as DCAT and ODRL for data models and access control) to provide easy to use information about the dataset for the different stakeholders. These stakeholders can include data providers (to find and manage their own data), data users, applications (machine readable formats), intermediaries and needs for governance. All stakeholders must have mechanisms to access and maintain relevant offerings in the data space, including creation, validation and updates of the data, services and offering.

DS2 view: Similar to several other building blocks, DS2 will address both the intra and inter dataspace scenarios that are unlikely to be significantly different. The metadata, quality, and policies must be in a level required by typical data space standards – need for usable data. The offering descriptions need to be defined targeting that the data can be used in the client data space.

3.6.2 Publication and discovery

Publication and discovery focus on the ability for data providers to make their data assets visible to other users in a data space. The publishing side requires ability to publish, modify and delete a parties own data sets including defined metadata (technical and commercial, e.g., what the data is as well as e.g., how it is priced). On the discovery side are multiple technologies and methods to improve the findability of data such as filters, trees, search (keyword, metadata, semantic descriptions etc.).

DS2 view: DS2 must address how data providers may want to make their data offering available beyond an individual data space. There may be separate selections (binary, dedicated metadata, dedicated modules) to allow publishing outside a participants own data space. Publication and Discovery must consider Data sovereignty since providers must understand who can see and fetch their data, after all they might have initiated a data space to work only with trusted partners and the governance may state free visibility to all data, which possesses a risk when opening to the outside world. On discovery side, consideration should also be given to the scalability of the identified solution as well as to the user journey.

3.6.3 Value added services

It is recognized that not everything can be specified in advance and different industries have domain specific requirements in their data spaces. Value added services (VAS) is a loose definition of additional modules in a data space targeting to fill gaps the generic framework does not answer. While the description of the value added services is open and no exclusive list of services included can be created, the specification lists examples of such services including data visualisation, data quality, anonymisation, marketplace, enrichment, compliance, maintenance, selection, extraction, combination, processing, transformation, different kind of AI services, innovation labs, links to different technologies, load balancing etc. The architecture of a data space should be modular enough that the VAS can be adopted easily in any data space implementation.

DS2 view: Some DS2 functionalities will fall into this category. The modules that will be classified as VAS will need to comply with the architecture principles of VAS on discovery, installation, maintenance, and administration. There is no common specification of a VAS installation, and DS2 should align DS2 specific VAS installations with single data space VAS installation schemes emerging during the project timeline. If a DS2 module is deemed to belong to VAS and other categories, it will be required to respect the architecture principles of both. Being the VAS world rapidly evolving and not fully specified, DS2 will need to constantly check possible evolutions in the IDSA specification to remain aligned with it. Finally, considering the important of VAS-type modules in DS2, it is important to validate each of them in at least one of the use cases.

4 USE CASE DESCRIPTIONS IN DS2

To see how the DS2 project can address the inter data space scenario, the use cases need to be analysed in more detail which will be presented in more detail in. Based on initial introduction of the use cases, DS2 proposes a data space scenario for the DS2 use cases. From each use case, a table of key information such as targets, high level data sources and key expectations for collaboration in the project, are presented. The use cases for the DS2 project are:

- Use Case 1. City Scape
- Use Case 2. Green Deal
- Use Case 3. Precision Agriculture
- Use Case 4: Inter Data Space Use Case.

The use cases will provide the DS2 project with data spaces for developing and testing the different functionalities developed in the project. An initial set of data spaces in the use cases is illustrated in figure 9.

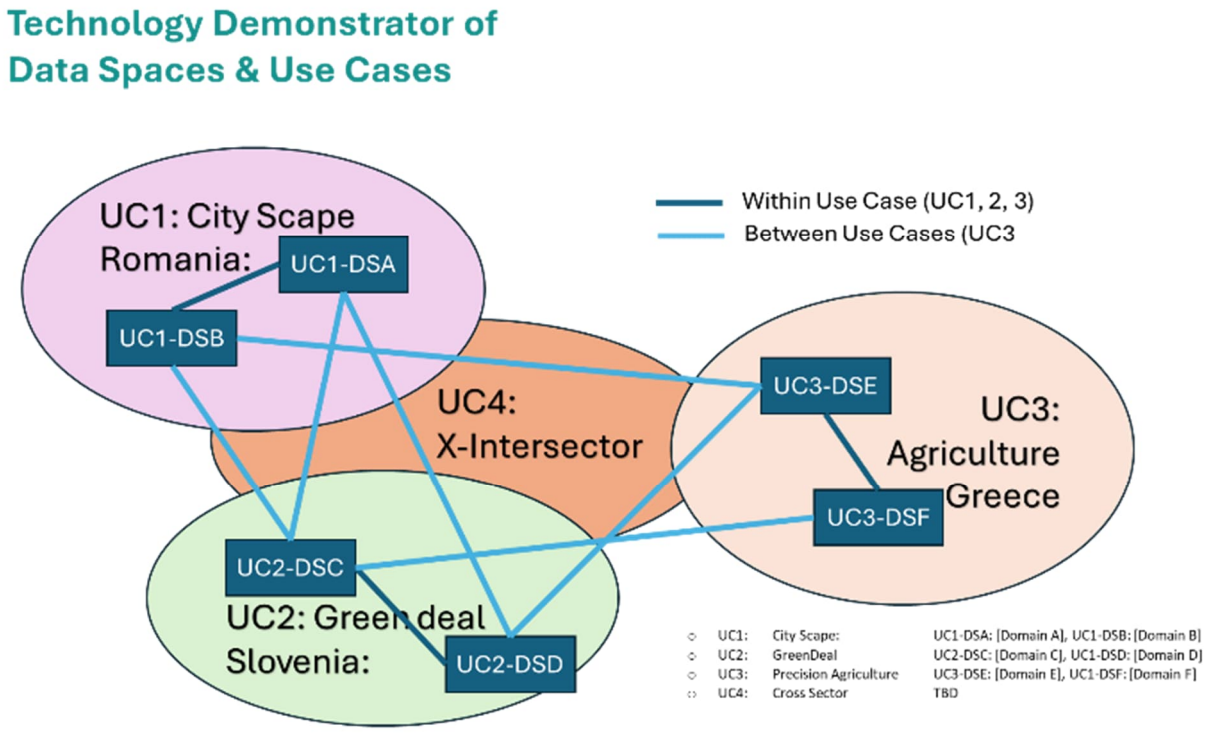

Figure 9. Different Data Spaces in Use Cases

4.1 Use Case 1: City Scape

Partners: CLUJ, ATIT, and DIGI

Demo site: Cluj Napoca (Romania)

Sectors: public administration, citizens, transport, energy, food supply chain, waste management

Targets of the Use Case	*Data of the Use Case*	*Expectations for Collaboration*
• Tangible benefits to the city of Cluj Napoca and its 420,000+ residents in achieving the Net Zero City objectives • Optimization of energy consumption and reduction of emissions • Increase the use and awareness regarding green affordable solutions in the residential and public building sectors • Facilitate collaboration and innovation, including the development of digital-twin solutions • Operationalization of policy commitments to open data sources and interoperability through concrete available tech solutions • Increased awareness on the use of data from the end-users	• Food supply chain, food loss and waste data using H2020 ZeroW data space • Energy consumption data from public buildings • Traffic data from the CLUJ's traffic monitoring system • GHG emissions data from national statistics portal.	• Governance • Practices • Policies • What can be learned from one another

Table 3. Description of Use Case 1

4.2 Use Case 2: Green Deal

Partners: MOMS, ITC, and DIGI

Demo site: Murska Sobota (Pomurje region, Slovenia)

Sectors: Agriculture, traffic, public administration, citizens, earth observation, and households

Targets of the Use Case	*Data of the Use Case*	*Expectations for Collaboration*
• Establish direct communication with residents, local stakeholders and policy makers by providing real-time data about air pollutants and GHG emissions • Awareness raising and operational/policy related strategies to be implemented and monitored.	• Food loss and waste data from H2020 ZeroW data space • Agriculture value chain data from DIVINE data space • Earth observation data for Murska Sobota city from public datasets of European GNSS (Sentinel 2 – Indexes relevant to agriculture, Sentinel 3 – GHG pollutants)	• Governance • Practices • Policies • What can be learned from one another

• To enable the municipality to make better decisions, policies, legislation • Being able to give information on air quality for kindergartens for example	• Traffic & weather data from national weather agency and data portals • Energy consumption (heating) data from public/private buildings of the city administration and additional energy consumption related data from the HE Green.Dat.AI data space.	

Table 4. Description of Use Case 2

4.3 Use Case 3: Precision Agriculture

Partners: CROWN, UTH, i4RI

Demo site: Greece

Sectors: Agriculture, weather, earth observation

Targets of the Use Case	Data of the Use Case	Expectations for Collaboration
• Faster training through Federated Learning, data collection from three different sectors, using more recent data and leading to more frequently updated models • Greater model accuracy and more training data from adjacent fields, more localized models, better reflecting local conditions by sharing data that the owners permit, and greater security via avoidance of data transmission to the core cloud, and increased awareness on the use of data from the end-users • To be able to provide agriculture consultancy for farmers • To enable farmers to share data between each other.	• Earth observation data from Sentinel 2 – Indexes relevant to agriculture, Sentinel 3 – pollutants) • Weather data (e.g. sunshine periods, rainfall) from national weather agency and UTH • CROWN's own IoT infrastructure for fertilizers use • Amount of water from irrigation systems • Soil quality, crop growth monitoring using cameras • Wind direction/speed sensors • Soil temperature/humidity	• Satellite data and nearby weather stations are interesting for UC1, yet they have not identified is this collaboration between different data spaces or data sources or something else. • Connecting with DS2 data spaces they are interested in demonstration purposes, however they see that it will be challenging

Table 5. Description of Use Case 3

4.4 Use case 4: Inter Data Space Use Case

As per the grant agreement, use case 4 is purposely not yet defined and an initial picture will emerge in deliverable D2.2 and serves to ensure that all features of DS2 can be demonstrated. The precise set of the technologies (The IDT toolkit and the applicable DS2 modules) will be determined when the data spaces in the use cases develop further. It is intention of the project to demonstrate interoperability of technologies between different data spaces based which are based on different initiatives (like IDSA and Gaia-X for example).

5 TECHNICAL PERSPECTIVE

In this chapter the current technological background of the DS2 technological project partners is presented with potential changes addressed as well. In addition, the relevant sister projects for DS2 are presented which were funded in the same call and the documents addresses the DS2 partners as liaisons for them.

5.1 Partner Knowledge and background update

5.1.1 From The Proposal

Samples of main technology/background brought in by technical partners and their TRL advancement are introduced. Potential changes to these technologies are addressed when applicable.

ICE: ICE Containerization used to facilitate multiplatform and modular deployment: Advance: Wizards to make Helm-charting process easier for developers: TRL 4 to 8.
CHANGES: No change. This will still provide the DS2 containerisation basis solely led by ICE and i4RI.

i4RI: i4RI is looking to exploit and enhance its knowledge in platform technology including a k8 modular platform used for integrating multiple modules inherited from the ZDMP project. TRL3-6.
CHANGES: As per proposal: i4RI will assist ICE with some aspect of containerisation, provide its marketplace background and support the module development process.

DIGI: DIGI's federated cloud environment will be extended in DS2 enabling collaboration and sharing of cross-sectoral, complex data among different organizations, communities, and other stakeholders with a shared interest. TRL 4 to 7.
CHANGES: No change.

SAG: SAG's enhancements for Cumulocity Cloud to extend the number of provided services for connectivity of networked digital devices and sensors. This will advance the capabilities and the applicability to a larger problem domain. TRL 5 to 8.
CHANGES: SAG will be using the open source thin-edge.io solution to push data from edge devices to cloud like Azure, AWS and its own Cumulocity IoT. Cumulocity IoT allows to write own micro services for connectivity of networked devices.

IBM: Parts of IBM's open source Fybrik framework for data security, which allows for policy-driven data access in a cloud environment. Fybrik was in development for over three years (TRL 6-7) and is part of a number of IBM offerings.
CHANGES: While IBM is not directly involved in trust and privacy management in DS2, it is expected that the concepts in Fybrik will be used in the development.

INDRA: INDRA's Onesait Platform provides the flexibility so that developers can build their own solutions in a solid and agile way using open source technologies, a flexible architecture and an innovative approach. Unique multi-repository data model. Collaborative environment for Data Scientist. Access management. TRL: 3 to 6.
CHANGES: No changes using Onesait Platform as technology provided and being developed/improved the modules indicated.

VTT: Data Space Innovation Lab (DSIL) provides experimental data space as a service. DS2 provides an excellent environment for its experimental deployment and further improvement of its capabilities and performance. TRL 3 to 6.
CHANGES: No changes

UOS: System Security Modeller Toolkit. (SSM): A cyber security risk management toolkit at TRL6 automating ISO27005 risk assessment process via a knowledge-based approach (https://zenodo.org/record/6676219). DS2 will extend its knowledge base from cybersecurity to account for complex data lifecycles and risks to provenance, sovereignty, privacy arising from sharing data in a data space. Knowledge TRL: 3 to 6.
CHANGES: No changes. The tool has also the name of Spyderisk.

ATC: ATC's blockchain-based system for data rights management. Advance: Via DS2 it will be extended to include a hybrid on-chain and off-chain system to store and manage the rights of digital assets. TRL 2 to 5.
ATC expertise in chatbot development. Advance: Through DS2, an AI-driven conversational assistant will be developed in the form of an intelligent 'chatbot', will be developed to efficiently elicit information related to complex DLC for data sharing between data spaces that will support data consumers and data providers. A similar AI-driven 'chatbot' will be developed via DS2 to intelligently moderate the efficient configuration of the complex DLCs, that will protect sovereign rights, including compliance with the EC regulations, over the complex life cycles of data sharing, aggregation and provenance. The AI-driven 'chatbot' will interface through the APIs of other WP3-6 modules to perform auto-assisted configuration based on the interaction and recommendations. TRL 3 to 7.
CHANGES: Regarding Task 3.3, TRL is updated to TRL 2-7.

INTU: INTU IntuScan "Meaning Mining" semantic text analytics platform is a mature patented system that performs ontologization of free text, converting the unstructured and polysemic natural language into structured normalized monosemic data that can be compared and matched with other data. It will be enhanced with additional features necessary for DS, so product TRL will not change but feature change will go from TRL 0 to 6.
CHANGES: Necessary changes for ontologization of structured metadata from different dataspaces. Adaptation of the database of output to the DB used by the project.

5.1.2 From New Activity

These are added following initial partner analysis and work allocation:

DIGI: DIGI's federated cloud environment is a part of the overall cloud based, secure Paradise platform. DIGI plans to integrate distributed data stores and Apache Kafka based data pipelines for efficient data transfer.

IBM: DIGI will contribute with its implementations of Oauth2.0 and/or OpenID Connect implementations for authentication and authorisation of actors involved in the edge-to-cloud enablement.

VTT: VTT has been developing in DSSC the Data Space Blueprint that is a repository of specifications and building blocks that can be used to create dataspaces. VTT is has had responsibility on Data Spaces Synergies work that is the foundation of inter data space interoperability solutions. VTT will bring these insights into DS2.

UOS: The risk assessment of data sharing across data spaces will include the sharing of Cyber Essential[7] information between data provider and consumer that will inform the decision of sharing data. UOS will work on the SSM knowledge base to improve the textual descriptions, generalising it, extracting, and encoding it in a format suitable to be used in LLM systems. Such corpus will be then linked to the training corpus used by ATC to train the chatbot. The result will be the ability of the chatbot to include in its knowledge risk management concepts and mitigation strategies, so presenting the user with a holistic approach at Dataspace deployment and configuration.

5.2 Final Expected Base Background Technologies

Based on the previous section, these are the technologies expected to be the basis for each task as background/knowledge although these may evolve as the deliverable D2.2: "Requirements, baselines, KPIs, Architecture & Specifications" is developed, and actual implementation takes place. The technologies are presented in table 4. In the table, related tasks to the technologies are also detailed when applicable.

Task ID	Tasks Name	Partner	Knowledge	License
3.2	GDPR, Ethics & Policy Compliance Data regulation modelling	UoS	Risk assessment tool	Apache
3.3	Rights Management & Sovereignty	ATC	Blockchain – Hyperledger Fabric (HLF)	Apache 2.0
4.1	Intersector Data Orchestration, Governance and Lifecycle	ICE	ICE Orchestration	Apache 2.0
		INDRA	Policy enforcement	Apache 2.0
4.2	AI-Driven Detection	INDRA	IDE for algorithms	Apache 2.0
4.3	Curation, Portability, Interoperability, and Transformation	INDRA	Transformation	Apache 2.0
WP5 and WP6	[Several]	IntuView	NLP	Propriety IntuView
WP5 and WP6	[Several]	IntuView	Ontology and algorithms related to ontologization of unstructured text	Propriety IntuView

[7] https://www.ncsc.gov.uk/cyberessentials/overview

5.2	Data Space Discovery and Assessment Module	ATC	Fine-tuning LLMs to perform specific tasks	Proprietary ATC (Not open source)
5.3	Recommend and configure complex DLCs	ATC	Fine-tuning LLMs to perform specific tasks	Proprietary ATC (Not open source)
5.3	Recommend and configure complex DLCs	UOS	Spyderisk (Knowledge Base)	Apache
6.1	Multi-cloud Data Sharing Infrastructure and Containerisation	INDRA	Data marketplace	Apache 2.0
6.2	Edge to Cloud Data Enablement, Control, and Data Quality Detection	DIGI	Components for data security and privacy	Apache 2.0
		INDRA	Data inspector	Apache 2.0
6.1	DSIL dataspace implementations	VTT	Implementation of IDSA data space (broker, DAPS, CA, and connectors + adaptation of EDC data space protocol v2)	Both proprietary and open source

Table 6. Base background technologies of DS2 partners

In terms of license nature, and as per the description of action for the DS2 project and table "final expected technologies per task" above, the vast majority of DS2 will be opensource, however there are technologies which are either stated to be 'not-open source' or that partner is still in internal discussions. These include Software AG (related to Edge to Cloud Data Enablement, Control, and Data Quality Detection) and ATC (related to Data Space Discovery and Assessments module and Recommend & Configuration module) who may consider commercial licenses.

Any technology which is not listed and subsequently used must be open source unless there is impacted partner and TM agreement. Similarly, partners may not covert unilaterally to non-open source within the project without impacted partner and TM agreement. The prerequisites will ensure there is control and visibility of potential IPR and joint development/dependency issues.

5.3 New Sister Projects of Potential Relevance

DS2 emerged from the call "HORIZON-CL4-2023-DATA-01-02: Integration of data life cycle, architectures and standards for complex data cycles and/or human factors, language (AI, data and robotics partnership)" (RIA) conducted in 2023-Q2. Projects which were similarly sponsored and started in the same period are listed below. Since both they and DS2 have only just started and are getting their "feet under the table", it is difficult as yet to determine potential reuse and interaction although this is planned through the task 1.3 Dissemination

task which includes clustering and DS2 has already been the initiator of many contacts. Some possibilities are listed although at this stage only based on website information since there are no common partners across the project with DS2. A general DS contact point to follow these projects is 21C as Impact Lead.

Project CEDAR: *CEDAR – Common European Data Spaces and Robust AI for Transparent Public Governance* – project will create an inclusive, federated, and privacy-preserving Data Commons space, to establish a secure and reliable space for sharing information among stakeholders across multiple domains. CEDAR will ensure that more data becomes available in the economy and society while keeping companies and individuals who generate the data in full control in a trusted, secure, and transparent manner. The project is coordinated by the Centre for Research and Technology Hellas (CERTH), Greece.

Potential DS2 Relevance: Exchanging information between stakeholders from various fields and the governance around this.

DS2 lead: UOS

Project PLIADES: PLIADES – *AI-Enabled Data Lifecycles Optimization and Data Spaces Integration for Increased Efficiency and Interoperability* – the project researches into novel, AI-enabled tools to advance full data life cycles integration, both within and between data spaces. The project will research and develop a novel data integration framework that builds on key SoA architectures and extends them with a series of advanced elements that solve essential, complex, yet practical problems around data green creation and storage, ownership and discovery, as well as use, re- use and disposal, among diverse data spaces. The project is coordinated by the Centre for Research and Technology Hellas (CERTH), Greece.

Potential DS2 Relevance: The Green deal use case.

DS2 lead: SAG

Project CyclOps: CyclOps – *Automated end-to-end data life cycle management for FAIR data integration, processing and re-use* – project proposes a new framework for the governance and maintenance of the complete data lifecycle for large-scale volumes of data. CyclOps will enable organizations to seamlessly provide, cross and analyse machine- and human-generated data from and for data spaces, thus facilitating the provision of added-value services on top. The project is coordinated by NTT DATA, Spain.

Potential DS2 Relevance: Data sharing and exchange in data spaces.

DS2 lead: VTT

Project NOUS: NOUS – *A catalyst for EuropeaN ClOUd Services in the era of data spaces, high-performance and edge computing* – project aims to create an architecture of a European Cloud Service that allows computational and data storage resources to be used from edge devices as well as supercomputers, through the HPC network, and Quantum Computers to unlock unprecedented potential for businesses, research institutions, and individuals alike. NOUS is coordinated by the AIR Institute, Spain.

Potential DS2 Relevance: Edge computing and components.

DS2 lead: IBM

6 PROBLEM SPACE DEFINITION

This section introduces identified problems of the inter data space information exchange divided into logical sections. The DS2 toolkit and modules are then introduced and linked to these problems. The problem space can be observed from multiple viewpoints as some challenges or functionalities are mandatory, and some are optional. This problem space document does not specify how the problems are solved but triggers the discussion on the potential solutions as well as helps to focus the DS2 project to choose the problems worth solving.

An example of a problem is finding relevant data spaces. The data spaces are, by definition, closed groups created for a certain purpose between named organisations ("participants"). They are most often not advertised anywhere, nor are they technically findable by any search robots, unless the data space authority wants the data space to be found. Potential solutions to find data spaces include 1) centralised database of data spaces 2) publishing contact details of a data space for example in internet page 3) word of mouth. As it can be seen, even the three different solutions require completely different approaches for a technical solution. The first would require a trusted organisation (or many) creating a standardised method for publication. The second would likely create a finite set of approaches to connect to a data space, ranging from APIs to emails. The final one would likely house almost infinite ways of contacting the data spaces. All the methods (and possibly others) can also co-exist. The underlying problem still persists, any project, such as DS2, which connects data spaces, must define the targets including how the data spaces are found, and build the functionality to support the selected environment.

6.1 Identified inter data space challenges

Based on the data space technical challenges related to operations inside a data space (section 2), and technological building blocks (presented in section 3), a combination of challenges related to inter data space interactions are illustrated in figure 10 below. In the following sub-sections the problem areas are discussed, and some solutions presented.

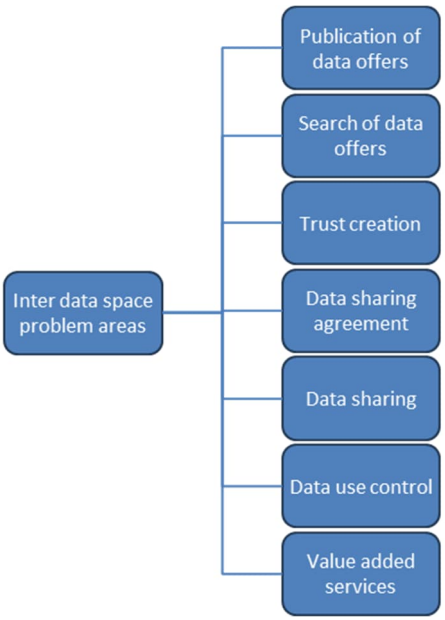

Figure 10. Problem areas for interconnecting data spaces

6.1.1 Publication of data offers

A data space is designed for data sharing between trusted participants within the data space. If access to the data is opened to a wider audience outside that data space, there must be mechanisms for data providers to control the visibility of the data offers. For example, a provider might want to have some offers visible only to their "own" data space participants whilst having some data offers visible to named other data spaces, or visible to all other (connected) data spaces, or even openly visible to everyone. This visibility must obviously follow the data space authority decisions on the data space visibility as well as a participants' own parameters.

6.1.2 Search for data offers

In addition to control the publishing of data offers, the participants of a data space need to be able to search for data offers in other data spaces. Inside one data space this is handled by the data catalogue component within the connector or as a service provided by the data space broker. When expanding the search to inter data space domain, the participants need to know exactly what is in a data space. Currently the existing community approaches, like the DSSC, have not yet identified a method for searching data offers between data spaces.

6.1.3 Trust creation

In the previous two points, data can be offered and data can be found, and the parties then have generic mechanisms to create the data sharing agreements between data spaces. The next step is to find and implement the trust anchors. The DSSC glossary defines trust anchors as "a party, a trust service provider or technical means which are accredited by the governance authority of the data space to be trustworthy anchors to digitally sign claims and/or attestations.[36]" First the stakeholders must evaluate the contract party, their infrastructure, and data quality to ensure that the other party can be trusted, the connectivity can be made, and the data will eventually be usable. Trustworthiness through agreed trust anchors can include for example:

- Identities
- Contracts
- Security
- Supervision (validation and law enforcement)
- Trust through common principle of data sovereignty
- Risk assessment

As the trust anchors are data space specific and unique, there will be problems when trust needs to be created between different data spaces using different trust anchors. Automating the creation of trust anchors to support inter data space connectivity might be a potential solution. For example, data space A may use a one identity provider, and data space B a different one leading to issues ranging from mutual recognition to the structure/provision of identity information.

6.1.4 Data Sharing Agreement

Dataspace or participants of data spaces must ensure compatibility of policies and exchange policy enforcement rules. There are different levels of policies already on a single data space level and between data spaces the levels (and actual policies) need to be agreed. The data space policies are defined in the figure 11 below.

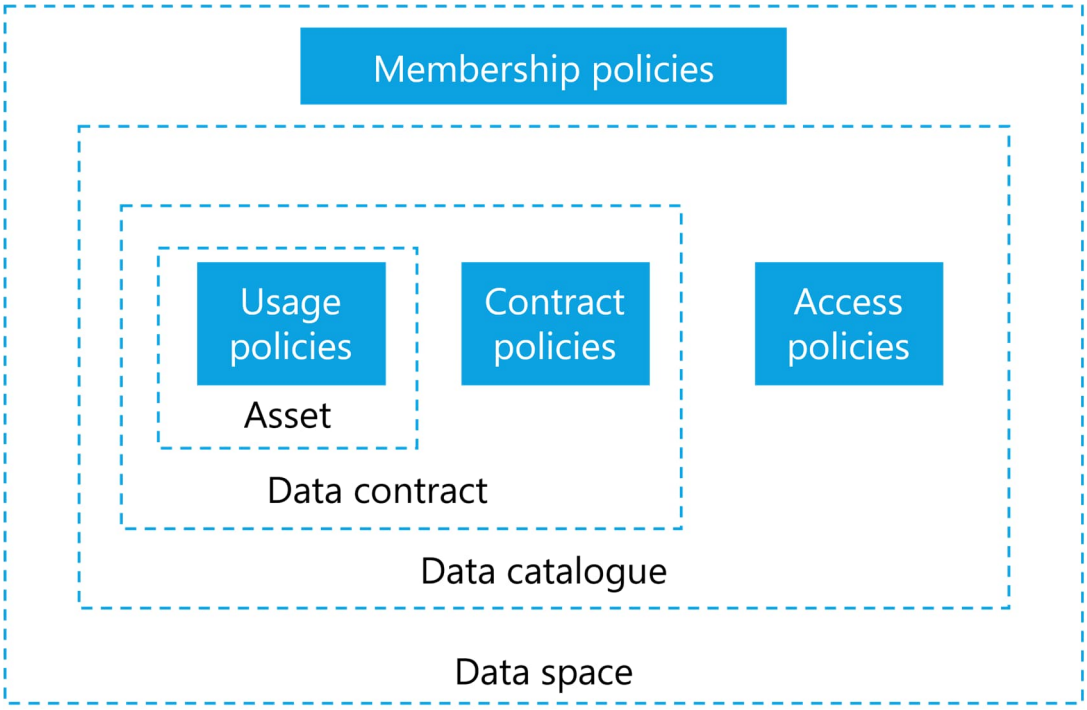

Figure 11. Different policies in a data space (based on DSSC [37])

It is likely all policies within a data space are relevant for inter data space use cases as well – hardly any data space would make inter data space connections with less policies than data space internal use. Some policies, such as access policies, might be stricter for inter data space connections, and some policies might need to be completely replaced such as the membership policies. There might also be a need for additional policies, e.g., membership policies on data space authority level and membership policies on individual organisation level (i.e., one policy for connecting the data spaces and another for a participant of data space to access another data space).

Expanding the operation between data spaces requires creation of agreements and policies between the data space governance authorities (who govern the data space). The data space authorities define the basic principles on how data is being shared from their data space. Also, the data space authority accepts the rules of other data spaces for data sharing. On-going community approaches have not yet established any mechanisms for inter data space agreements.

Single participants of a data space do not decide issues on a data space level and cannot establish inter data space agreements. Adding independent participants to a data space are a data space governance authority decision, also allowing whole data spaces in "join-in" in some level must be also a data space governance authority decision.

6.1.5 Data Sharing

On top of agreements between data spaces there must be contracts, preferably partially predefined, for the participants of the data spaces and an agreed process to making contracts between participants. The process of data sharing in an inter data space scenario should not be too complex since much of the work will have already been perform in a data space. Predefined contracts encourage taking the first steps in the path of connecting data spaces. Such predefined contracts can be guidelines, templates, strict contracts where no changes are allowed etc.

There should also be instructions or process to terminate contracts and agreements on both participant level and data space level. This part is also essential catalyst for creating the connections in the first place. Data space participants need to know that the contracts can be withdrawn in case of misuse or if benefits are lower than the risks. When policies between data spaces are different there will need to be a negotiation to find common or 'lowest common denominator' policies. When participants break such policies there will typically be arbitration or legal remedies to resolve this but whilst relevant to data space sharing is outside of the implementation scope of DS2.

6.1.6 Data Use Control

Data use control services are related to ensuring the usability of data and monitoring the use of data. There are two phases for data use control services: services before the data transfer and services after the data has been transferred.

Before the data transfer, ensuring the interoperability of data space communications include access control and mechanisms for data transfer. When all preparation tasks (discovery, authentication etc.) are done, the data exchange can take place. Depending on the selected data space connectivity method (e.g., federative, intermediary, aggregator, point-to-point) the data transfer might appear different although the underlying protocols would be identical. All functionalities related to data transfer must be able to fulfil the governance principles of both data spaces involved. One example of such functionality is logging the transactions between different data space participants. In inter data space connections logging all users of other data space to their own identity management system is not feasible, and other solutions need to be explored.

The provider of the data has already in the offering phase transformed the data into a data product that is perceived to be of value for the other participants. However more action to ensure the quality of data can be performed after the transfer has been concluded. When the data has been transferred between different participants, there should be mechanisms for feedback about the quality and usability of the data. These mechanisms may be used by data space participants to further analyse quality, these mechanisms are initially meant for intra data space usage. How to ensure the compatibility of these services in inter data space transfers is unexplored. For example, there can be multiple services transforming the data into a form that a data consumer can eventually use for its business. Some features in section 6.2 are designed to improve the usability.

6.1.7 Value added services

Value added services can be considered important channels for providing some of the key functionalities for inter data space interactivity. Value add services were introduced in section 2, figure 1. These services provide functionalities for different data space participants, examples include [38]

- data services (E.g., data transformation services and processing services)
- use case specific services (E.g., collaborative analytics services and marketplaces)
- application services (E.g., AI applications)
- infrastructure services (E.g., service orchestrator and multicloud environment services)

Currently these services are considered for in data space use, but they can also offer an opportunity for services enabling inter data space functionalities. For the DS2 project, some of the functionalities could include combining different ontologies or performing data transformations between data spaces.

6.2 Addressing the challenges with the DS2 toolkit functionalities

6.2.1 Problem Spaces and DS2 functionalities

The DS2 project will provide a toolkit (Intersector DataSpace Toolkit) which is consists of a broker and a set of plug-in modules to facilitate data sharing oriented operations. With the modules, the DS2 project will help different data space participants in promoting and searching for data offers, creating trust and agreements and sharing data between data spaces. The set of modules and tools are (based on the description of action):

- DS2 Sovereignty Decision Support Tool: Tool for sovereignty decision support and regulatory guidance for practitioners
- DS2 Risk Modeller Module: Tool for risk analysis
- DS2 Blockchain DRM Tool: A blockchain-based Digital asset Rights Management (DRM) system, enabling flexible control over shared data and non-repudiable audit trails of access to it.
- DS2 Orchestration Module: Data Space Operators benefit from the ability to combine data from multiple domains/sectors to gain better insights
- DS2 AI Detection Module: AI driven anomaly detection to enhance quality of the data
- DS2 Curation Module: Organises and integrates complex data collected from various sources and sectors
- DS2 Portability Module: A plugin that securely exchanges data from one platform/service to another (e.g., data owner to DS2 service)
- DS2 Interoperability and Transformation Module: Performs any transformation due to data sovereignty requirements, ontological reconciliation, and schema differences leading to an interoperable data format
- DS2 Culture and Language Module: Contains an ontology and a software toolkit for multi-cultural and multi-lingual data services
- DS2 Data Space Discover and Assess Module: Supports data owners in discovery and assessment of the technical, business, operational and organisational capabilities data spaces complex data sharing
- DS2 Recommend and Configure Module: Supports data consumer and data providers with recommendations and configuration of data life cycle covering provenance, sharing, aggregation. It also takes care of sovereign rights and compliance with regulations
- DS2 Containerisation Module: Packages data and applications into portable entities, facilitating deployments, portability and management federated data infrastructure
- DS2 E2C Module: Enables an edge to cloud connectivity through apps and devices capable of collecting, processing data and interconnecting this data with the cloud infrastructure
- DS2 Data Quality Module: A tool defining and constantly monitoring complex data quality and establishes a framework for automatic checks, e.g., of loss detection

The different modules provide functionalities for data sharing. Figure 12 illustrates an initial breakdown of different DS2 toolkit functionalities provided by the modules introduced earlier. These toolkit functionalities are then matched to inter data space challenges they can help to alleviate.

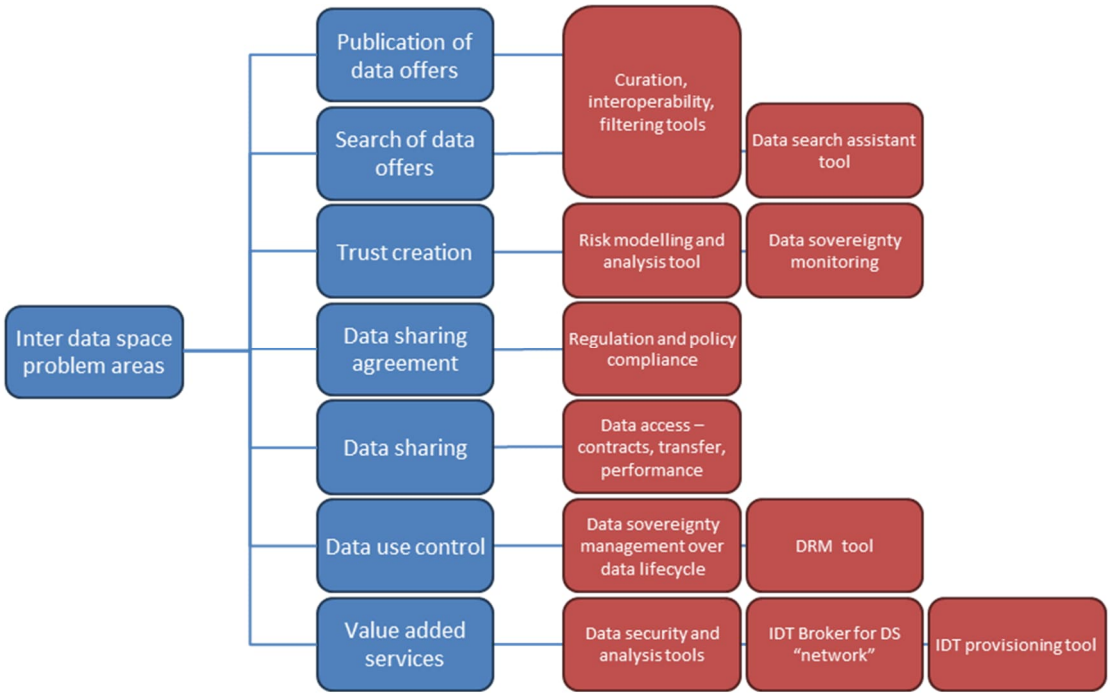

Figure 12. DS2 toolkit functionalities matched to inter data space challenges.

Targets of the use case have been presented in section 4. The use cases have expressed that they will benefit from the following initial set of DS2 functionalities and this will be firmed up in D2.2:

- Data Space Discover and Assess Module
- Recommend and Configure Module
- Interoperability Module
- Transformation Module
- Recommend and Configure Module
- Risk assessment

The modules and tools are aligned into different tiers:

- Tier 0: Supporting data spaces
- Tier 1: Services related to finding and deploying DS2 modules before operational phase of data sharing
- Tier 2: Enabling intra data space sharing
- Tier 3: Enabling inter data space sharing

The DS2 modules and tools are allocated into different tiers as presented in table 7 below:

Tier	Module/Feature Outcome	Initial Modules
Tier 3: DS2 Inter Data Space Sharing	Publication of data offers	DS2 Offer Publication Module
	Search of data offers	DS2 Discover and Assess Module

	Trust Creation	DS2 Risk Modeller Module
		DS2 Data Sovereignty Decision Support Tool
	Data Sharing Agreement	DS2 Trust Module
	Data Sharing	DS2 Blockchain DRM Tool
		DS2 Interoperability Module
	Data Use Control	DS2 Blockchain DRM Tool
		DS2 Sovereignty Decision Support Tool
Tier 2: DS2 In data space enablement	Value Added Services	DS2 Orchestration Module
		DS2 AI Detection Module
		DS2 Curation Module
		DS2 Data Quality Module
		DS2 Portability Module
		DS2 E2C Module
Tier 1: DS2 Marketplace and deployment		DS2 IDT Broker
		DS2 Portal
		DS2 Containerisation Module
Tier 0: DS2 Support		DS2 Culture and Language Module
		DS2 Identification Module

Table 7. DS2 Modules and tools allocated into different tier levels

6.2.2 Functionality of DS2 functionalities in data spaces

FFigure 13 below was presented in the section 2.2 earlier and it defines the positioning of data management and data spaces in a multi data space environment.

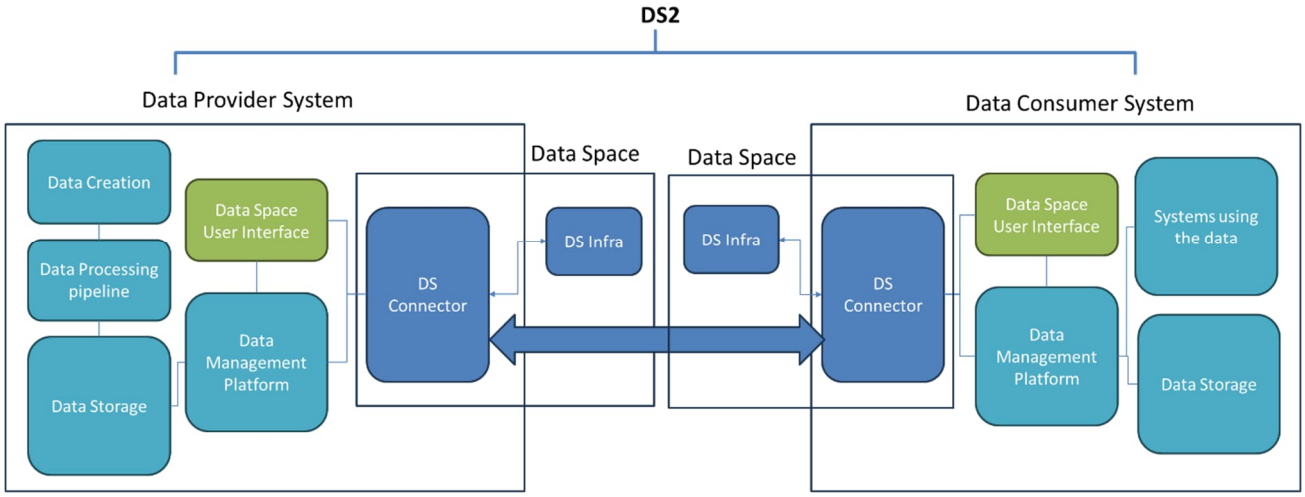

Figure 13. Data pipelines over data spaces

There are five places in the model where any such modules could operate:

- In DS infrastructure
- In DS connector
- Outside a data space but within data space participant's data management system
- Within the DS2 data sharing environment
- Outside all systems presented in the figure

The different DS2 modules should by default operate in the DS infrastructure or within the connector (see elements in in figure 13). In some special cases, the modules may also be deployed in other parts of the data pipeline if feasible and exceptionally some modules can be deployed outside the picture, for example in app-store like environment.

7 CONCLUSIONS

This deliverable defines what is a data space, explores different data space approaches and their relevance for the DS2 project. The DS2 project is then matched against the DSSC building blocks to analyse the future requirements for the DS2 functionalities. The deliverable also introduces the use cases of DS2 and preliminary matches them with the DS2 functionalities. The general problem space for the DS2 project is then presented. The deliverable also introduces some regulatory issues relevant for DS2 as well as emerging technological standards that require attention during the project.

This deliverable has introduced some of the key challenges related to inter data space interactions and explored how the DS2 modules and the IDT toolkit can address these challenges. The DS2 modules have potential to address many of the inter data space and dataspace enabling challenges. The IDT toolkit and the DS2 modules have a potential to ease the deployment of data spaces and sharing of data between the data spaces. However, not all challenges related to inter data space usage are addressable by the DS2 project.

Another important avenue for impact is for example through the multiple different approaches and projects related data spaces, where the DS2 project partners are active, such as IDSA or the DSSC project. Additional developments are needed within the concept of a data space and functionalities within a data space to enable and support interaction between different data spaces. For the DS2 project, close cooperation related to initiatives such as the DSSC is important to further develop the concept of data spaces towards more usable concept for building up the future of data spaces.

ANNEX A: REFERENCES

[1] A. Hutterer, B. Krumay, and M. Mühlburger, 'What Constitutes a Data space? Conceptual Clarity beyond Technical Aspects', 2003.

[2] Common European Data Spaces | Shaping Europe's digital future'. Accessed: Jun. 20, 2024. [Online]. Available: https://digital-strategy.ec.europa.eu/en/policies/data-spaces

[3] B. Otto and M. Jarke, 'Designing a multi-sided data platform: findings from the International Data Spaces case', *Electronic Markets*, vol. 29, 28 2019, doi: 10.1007/s12525-019-00362-x.

[4] Poikola, Antti, Verdonck, Bert, Joosten, Rieks, Guggenberger, Tobias, and Salminen, Satu, 'DSSC Glossary | Version 2.0 | September 2023 - Glossary - Data Spaces Support Centre'. Accessed: May 14, 2024. [Online]. Available: https://dssc.eu/space/Glossary/176553985/DSSC+Glossary+%7C+Version+2.0+%7C+September+2023

[5] G. Giussani, S. Steinbuss, M. Holesch, and N. Gras, 'Data Connector Report', [object Object], Mar. 2024. doi: 10.5281/ZENODO.10869525.

[6] Data space use cases and business model - Glossary - Data Spaces Support Centre'. Accessed: May 27, 2024. [Online]. Available: https://dssc.eu/space/Glossary/176554067/3.+Data+space+use+cases+and+business+model

[7] E. Curry, S. Scerri, and T. Tuikka, Eds., *Data Spaces: Design, Deployment and Future Directions*. Springer Nature, 2022. doi: 10.1007/978-3-030-98636-0.

[8] A. Poikola, Laszkowicz, P. J., Takanen, Ville, and Toivonen, Teemu, 'The Technology Landscape of Data Spaces', Oct. 2023. [Online]. Available: https://media.sitra.fi/app/uploads/2023/10/sitra-technology-landscape-of-data-spaces.pdf

[9] 'IDS RAM 4.0 | IDS Knowledge Base'. Accessed: May 30, 2024. [Online]. Available: https://docs.internationaldataspaces.org/knowledge-base/ids-ram-4.0

[10] 'Dataspace Protocol 2024-1 | Dataspace Protocol | IDS Knowledge Base'. Accessed: May 30, 2024. [Online]. Available: https://docs.internationaldataspaces.org/ids-knowledgebase/v/dataspace-protocol/

[11] 'Introduction | IDSA Rulebook | IDS Knowledge Base'. Accessed: May 20, 2024. [Online]. Available: https://docs.internationaldataspaces.org/ids-knowledgebase/v/idsa-rulebook/idsa-rulebook/1_introduction

[12] M. I. S. Oliveira and B. F. Lóscio, 'What is a data ecosystem?', in *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*, Delft The Netherlands: ACM, May 2018, pp. 1–9. doi: 10.1145/3209281.3209335.

[13] 'Communication Material - Gaia-X: A Federated Secure Data Infrastructure'. Accessed: Jun. 26, 2024. [Online]. Available: https://gaia-x.eu/news-press/communication-material/

[14] 'Gaia-X Trust Framework - Gaia-X Trust Framework - main version (fb420580)'. Accessed: May 30, 2024. [Online]. Available: https://gaia-x.gitlab.io/policy-rules-committee/trust-framework/gaia-x_trust_framework/

[15] 'FIWARE-Ops/data-space-connector'. FIWARE Ops, Mar. 30, 2024. Accessed: May 28, 2024. [Online]. Available: https://github.com/FIWARE-Ops/data-space-connector

[16] 'Smart Data Models', Smart Data Models. Accessed: May 28, 2024. [Online]. Available: https://smartdatamodels.org/

[17] 'Introduction to Smart Data Models', Accessed: May 20, 2024. [Online]. Available: https://docs.google.com/presentation/d/1k021ZzRyk2PtikubOFNa3KyXeXL5nTmWWxyhs0Y0r1k

[18] 'Data Spaces Support Centre', Sitra. Accessed: May 28, 2024. [Online]. Available: https://www.sitra.fi/en/projects/data-spaces-support-centre/

[19] 'Data Spaces Support Centre'. Accessed: May 28, 2024. [Online]. Available: https://dssc.eu/space/Mission/175308804/Mission+and+Vision

[20] 'Simpl: Cloud-to-edge federations empowering EU data spaces | Shaping Europe's digital future'. Accessed: Mar. 22, 2024. [Online]. Available: https://digital-strategy.ec.europa.eu/en/policies/simpl

[21] 'AWS for Data Spaces', Amazon Web Services, Inc. Accessed: Jun. 20, 2024. [Online]. Available: https://aws.amazon.com/government-education/aws-for-data-spaces/

[22] 'Trusted Data Sharing Framework', Infocomm Media Development Authority. Accessed: May 27, 2024. [Online]. Available: https://www.imda.gov.sg/how-we-can-help/data-innovation/trusted-data-sharing-framework

[23] *IDSA Ecosystem Building Call Open Month | Edition 1 | May 6, 2024*, (May 08, 2024). Accessed: Jun. 20, 2024. [Online Video]. Available: https://www.youtube.com/watch?v=4aegAjs9q6U

[24] 'Google Cloud joins Catena-X to build sovereign data ecosystem', Google Cloud Blog. Accessed: Jun. 20, 2024. [Online]. Available: https://cloud.google.com/blog/topics/manufacturing/google-cloud-joins-catena-x-to-build-sovereign-data-ecosystem

[25] Data Space Support Center, 'Collection of Standards and Technologies landscape | Version 1.0 | October 2023 - Collection of Standards - Data Spaces Support Centre'. Accessed: Apr. 25, 2024. [Online]. Available: https://dssc.eu/space/SE1/185794561/Collection+of+Standards+and+Technologies+landscape+%7C+Version+1.0+%7C+October+2023

[26] N. Gras, 'New EU regulatory environment for data spaces', International Data Spaces. Accessed: Mar. 22, 2024. [Online]. Available: https://internationaldataspaces.org/new-eu-regulatory-environment-for-data-spaces/

[27] S. J. Hilberg and S. Steinbuss, 'Reflections on the DGA and Data Intermediaries', Zenodo, Oct. 2023. doi: 10.5281/ZENODO.10117605.

[28] ‘Data Act | Shaping Europe’s digital future’. Accessed: Mar. 22, 2024. [Online]. Available: https://digital-strategy.ec.europa.eu/en/policies/data-act

[29] ‘Data Act | Shaping Europe’s digital future’. Accessed: May 27, 2024. [Online]. Available: https://digital-strategy.ec.europa.eu/en/policies/data-act

[30] ‘FAQ’, GDPR.eu. Accessed: May 27, 2024. [Online]. Available: https://gdpr.eu/faq/

[31] ‘Regulatory Compliance - Blueprint v1.0 - Data Spaces Support Centre’. Accessed: May 27, 2024. [Online]. Available: https://dssc.eu/space/BVE/357074696/Regulatory+Compliance

[32] IDSA, ‘Widening the scope: the anatomy of the Data Spaces Radar’. Accessed: May 30, 2024. [Online]. Available: https://www.dataspaces-radar.org/

[33] T. de Bruin *et al.*, ‘Do I-PASS for FAIR. A self assessment tool to measure the FAIR-ness of an organization’, Nov. 2020, doi: 10.5281/zenodo.4080867.

[34] ‘Building Blocks | Version 0.5 | September 2023 - Building Blocks - Data Spaces Support Centre’. Accessed: May 30, 2024. [Online]. Available: https://dssc.eu/space/BBE/178421761/Building+Blocks+%7C+Version+0.5+%7C+September+2023

[35] ‘DS4Skills Blueprint’. Accessed: Jun. 20, 2024. [Online]. Available: https://skillsdataspace-blueprint.eu/