INTERNATIONAL DATA SPACES ASSOCIATION

# Please note that front and back covers will be inserted into the PDF file. The intro text will appear here.

Dummy text to be replaced. The position paper underlines the importance of data spaces and though the sovereign sharing of data in creating the future data economy. It has been developed under the coordination and leadership of Task Force 1 lead by International Data Spaces Association of the Horizon 2020 project "OPEN DEI Aligning Reference Architectures, Open Platforms and Large-Scale Pilots in Digitising European Industry" with the collaboration of more than 40 data spaces and industrial domain experts representing more than 25 organisations from 13 Horizon 2020 projects and related initiatives. This is the first approach to define the design principles for data spaces, agreements on the building blocks for a soft infrastructure and governance for data spaces.

## Contributing Organizations

**< Logos >**

## Contributing Projects

**< Logos >**

INTERNATIONAL DATA SPACES ASSOCIATION

**Publisher**

International Data Spaces Association
Anna-Louisa-Karsch-Str. 2
10178 Berlin
Germany

**Copyright**

International Data Spaces Association, Dortmund 2021

**Editor**

< Names >

**Authors & Contributors**

*Dummy text to be replaced. This paper has received funding from the European Horizon 2020 Programme for research, technological development and demonstration under grant agreement n° 857065*

# Table of content

# List of figures

# 1 Introduction

## 1.1 Who should read this rulebook?

It is all about data. If you are using data-driven ecosystems or data-driven business models, you should build or join a data space – and therefore read this rulebook. And if you are not yet, think about it!

Data sharing is becoming a critical success factor for all businesses and organizations in all national and international economies. Data access and sharing also helps in meeting specific societal, policy, and legal objectives that are in the public interest. This rulebook covers several types of data sharing: data sharing ecosystems, peer-to-peer data sharing, data marketplaces and data-driven platforms.

The data space approach described in this rulebook is for anybody interested in trusted and secure data access and sharing. It is relevant to businesses, organizations and individuals wanting to learn how their data rights can be handled in these data spaces.

## 1.2 Goals and scope of the IDSA rulebook

### 1.2.1 Goals of the IDSA

The International Data Spaces Association (IDSA) has defined a data sharing scheme (IDS), including a reference architecture, open source building blocks, and a certification process for creating and operating data spaces. IDS is based on commonly accepted data governance models facilitating secure sharing and easy linking of data within business ecosystems. The goal of IDSA is to make IDS a global standard for sovereign data sharing.

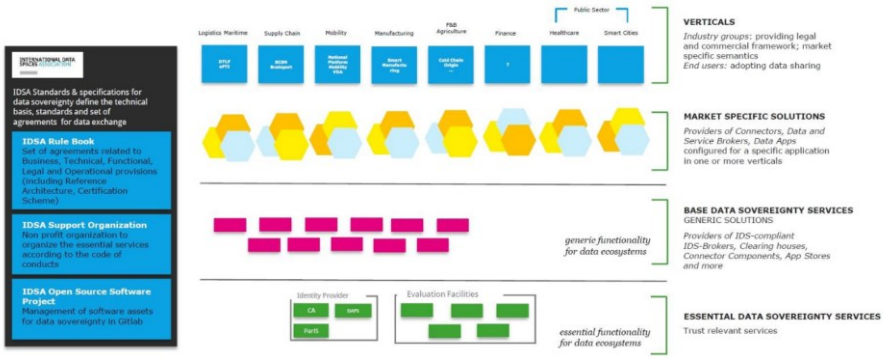



*Figure 1 Overview IDS enabled ecosystems*

**Commented [TM1]:** This viewgraph may need to be revised.

The most important design principle for data spaces is to ensure data sovereignty for all data. This even enables the sharing of sensitive and most valuable data assets between selected participants. The IDS scheme guarantees data sovereignty for data owners who provide the shared data. This is the basis for offering smart services and for establishing innovative business processes.

IDSA defines the technical foundation and a set of agreements for secure and trusted data spaces, where companies of all sizes and industries can manage their data assets. The association already counts over one hundred and thirty member organizations from twenty countries. The interplay of all these organizations as data space participants and service providers will deliver on the shared value proposition of generating business value from data.

### 1.2.2 The purpose and scope of the rulebook

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:

- Guidelines for the functionality of common services are presented as well as the definition, processes, and services of specific roles.
- Guidelines how to implement or use a technical artefact of the IDSA.
- Guidelines for the work and collaboration within data services.
- Guidelines for the legal basis in compliance with the regulatory environment to ensure trust and security.

This framework applies to all IDS-related roles and their interaction in the specific environment:

1. The IDSA support organization is responsible for maintaining this rulebook and supports its application. It enables the orchestration of processes and the realization of interfaces to other parties.

2. The essential service providers make these services available to the participants. They are the source of common agreements.

3. All IDS users are getting guidance on how to proceed in realizing use cases based on a trustworthy infrastructure and governance.

Commented [SS2]: To be updated

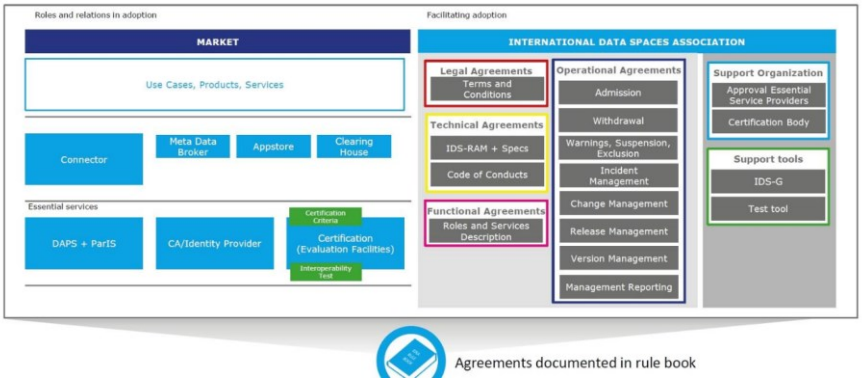



*Figure 2 Overview Rule Book scope and goals*

## 1.3 Relationship with other initiatives

Commented [SS3]: Lars Nagel has a good figure this

Commented [TM4R3]: I found this picture from Lars that could be fitting?

Commented [AS(5R3]: I think this chapter about the other initiatives could go towards the end of the rulebook. Next to outlook. Since a lot is introduced in this rulebook I would stay on message and not introduce too much (and distract) so early.

### 1.3.1 Data Spaces Business Alliance (DSBA)

Four key European organizations (IDSA, Gaia-X, FIWARE, BDVA/DAIRO) have formed an alliance creating one voice and a common framework to make data spaces happen. Together, the DSBA represents 1,000+ leading industry players, associations, research organizations, innovators, and policymakers worldwide. With its combined cross-industry expertise, resources, and know-how, the DSBA drives awareness and technology adoption, shapes standards and enables integration of data spaces across industries.

The aim of IDSA rulebook is to ensure compatibility with the common data space framework as envisioned and defined by the DSBA.

### 1.3.2 The Data Space Landscape

Further information on the landscape of data spaces are subject of a dedicated IDSA publication, the Data Space Landscape[to be published end of February 2023].

## 1.4 Related documents

You can find additional information about IDS related elements at other sources provided by IDSA:

- The IDSA website (https://www.internationaldataspaces.org) reflects what we do, who we are and what International Data Spaces stands for. Use cases illustrate the possibilities of the data economy and outline the added value created by the IDS standard. The download center gives access to the IDS Reference Architecture, papers and studies, scientific publications, and marketing material. Our content is constantly updated with news, blog articles, events and our regularly published magazine DATA SPACES NOW!

- The IDSA GitHub repositories (https://github.com/International-Data-Spaces-Association) see also section 3.4

# 2 Guiding principles

The IDSA rulebook is based on a set of generic principles and underlying values. The key aspects are related to the governance of data spaces and the roles actors can have.

**Not reinventing the wheel:** use proven technologies

**Integrate existing systems:** integrate data spaces into existing systems to the extent possible

**Integrate or use existing standards**: align national and international specifications, technical standards, and established processes

**Industry and domain independent**: make data spaces applicable as a concept as a horizontal standard

**Easy to use**: low deployment threshold for companies and initiatives with a focus on portability and replicability

IDSA applies four key governance principles: accountability, transparency, fairness, and responsibility. As a result, IDSA offers free use of IDS specifications and related open resources for all, open governance processes in which everyone can participate, transparent decision making - preferably by consensus.

## 2.1 Overarching considerations of data spaces

### 2.1.1 Introduction

Data and technology – and also data spaces – are both: *never* neutral and *always* neutral. They are never neutral in the sense that they are always parts of complex, human systems which reflect the values of the people involved. Data sets are collected by people, who decide what data to collect and how. These choices, in turn, are linked to values, they indicate what data people consider important to measure and collect.

Data and technology are also always neutral in the sense that they can be used for purposes that support or go against the values of their users and their societies. A classic example of this is nuclear technology, which gave us both the atomic bomb and radiation therapy to treat cancer.

To identify these aspects for data spaces we use PESTLE analysis - a tool to describe a macro picture of the environment of a data space. PESTLE stands for **p**olitical, **e**conomic, **s**ocial, **t**echnical, **l**egal and **e**nvironmental. For each section, we *describe* the (European) values embedded in IDS-compliant data spaces and do *not prescribe* specific purposes for which these data spaces may be used. This allows users of this rulebook to critically reflect the values embedded in their own data space.

Solid values and ethics are fundamental to any technical implementation; their absence has led to catastrophic effects on humanity. The use of data needs good governance goals. We

are deeply rooted in the European values of freedom, inviolability, privacy, security, humanity, and respect (without claiming to be exhaustive) and therefore include considerations of values and ethics into the rulebook, and carefully choose the path to the data economy weighing the impact on people and societies.

**P** Political

The political perspective in the European Union

Data sharing and data sovereignty are at the core of the European Data Strategy[11] (2020). Recognizing that industrial and commercial data are key drivers of the digital economy, the strategy uses "sovereignty" to describe its ambition to keep control of data with those who generate it.

Data spaces are an important means to strengthen ddigital sovereignty - a cornerstone of the European Digital Decade proposal[9] as highlighted by EC President Ursula von der Leyen's State of the Union Address to the European Parliament in 2020[10]. As outlined in the European Data Strategy, data spaces will empower data users and data holders to establish a healthy balance between the rights and interests of all stakeholders involved, with the objective of a wide use of data. Data spaces will empower data users and data holders to establish a healthy balance between the rights and interests of all stakeholders involved. This is outlined in the European Data Strategy - with the objective of a wide use of data.

The European Commission's policy proposal "Path to the Digital Decade" aims for a digital transformation of the Union by 2030. T addressing the challenges and ambition objectives are described in the Commission's "2030 Digital Compass"[12]. The Commission proposes sSeveral legislative instruments are being proposed by the European Commission for theto implementation of the European Data Strategy, notably: i) the Data Governance Act (DGA, Nov 2020) with a focus on ensuring trust in data transactions, ii) the Digital Markets Act (DMA, Dec 2020) regulating data based market power; iii) the AI Act (2021) with implications foron AI data governance and data management; iv) the Implementing Act on high-value data sets under the Open Data Directive to further unlock the socio-economic potential of data as a public good, and v) the Data Act (DA, Feb 2022) targeting a wide spectrum of topics, including facilitating access to and use of data by businesses and consumers, and enabling public sector bodies and institutions to use data held by enterprises in exceptional situationscircumstances.

Challenges stem from the complexity of the legal framework (EU vs. national, horizontal vs. sector-specific, economic law vs. fundamental rights, etc.) and competing relationships between stakeholders in data spaces. This highlights the need for legal interoperability: a common understanding of the evolving legal environment that is evolving, a; common vocabulary (legal-technical) and; facilitating the implementation of the balance between policy objectives. The realization of data spaces requires policies that can adapt to respective specificities and to their dynamic evolution over time, while aiming atfor a common European data space.

Finally, the EU emphasizes in the "An EU Strategy on Standardization setting global standards in support of a resilient, green and digital EU single market" the EU emphasizes the importance of the success of European actors in standardization at international level. It will strengthen Europe's competitiveness, technological sovereignty, and will also protect EU values. One of the priority areas identified is "data standards enhancing data interoperability, data sharing and data re-use in support of the Common European Data Spaces".

**Commented [VL6]:** I moved this to be a part of the above text.

**Commented [GC7]:** @Viivi @Marco now that the DSSC grant is signed, we can re-use the text from its proposal. Chapter 1.2 of the original document described the expected DSSC contribution to long-term EU policy objectives, policies and strategies. I've removed the references to DSSC, and the remaining text is a nice description of the EU political landscape.

What do you think?

We should give credit to BDVA's Ana Garcia that I seem was one of the main authors of this section

**Formatted:** Font color: Auto

## **E** Economic

The overarching goals for IDSA include making more data available to more organizations and ecosystems, recognizing that the availability and sharing of data is a critical success factor for local, national, and international economies.

Economic benefits happen in a data space at~~on~~ two levels: directly through sharing or accessing data that is of value to participants (micro-level: ego-system) and indirectly through supporting/creating a larger ecosystem that benefits all participants (macro-level, eco-system).

A digitally supported value chain can facilitate collaboration and improve resilience by identifying deviations or threats early (for example resource scarcity in a value chain). Access to even broader collaboration can ~~further~~ unlock potential when multiple data spaces are connected.

In terms of fairness, benefits can be spread throughout the value chain. Often large benefits can be achieved at a later stage at the expense of efforts at an earlier stage~~. Through consensual agreements in the data space this can be mutually beneficial~~. Consensual agreements in the data space can make this mutually beneficial.

## **S** Social

Commented [VL8]: I added this paragraph now, it's pretty high-level but hopefully useful & going in the right direction.

The social values embedded in the work of IDSA data spaces ~~work~~ are European ideals ~~including~~ such as freedom, inviolability, privacy, security, humanity, and respect. Issues such as ~~equity of~~ gender equality, socio-economic opportunity, and cultural representation are relevant wherever data is collected. Exactly *how* these values manifest in each data space is up to the implementer to decide - in collaboration with all stakeholders. The needs and priorities of specific economies, ecosystems, and communities vary. Our overarching societa~~a~~l value commitment is *pluralism* of *interoperable* and mutually *respectful* data spaces whose values and priorities are defined in an *inclusive* manner.

## **T** Technical

Data spaces should be built on widely established and openly accessible protocols, standards, and technical frameworks. Interoperability standards define the boundaries between two objects that have gone through a consensus process. The consensus process should have a narrow technical focus (like W3C, OASIS). W3C has developed processes and policies that promote the development of high-quality, consensus-based standards, many of which power

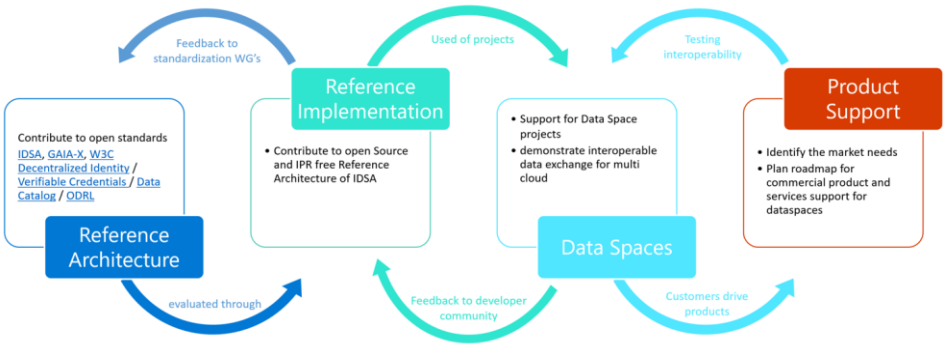



*Figure 3 Collaborative Development of Architectures and Implementations in Data Spaces*

the web and enterprise computing. ISO and IEC are adopting W3C technology and guidelines for a broad industry use.

When standards are adopted successfully, best practices show that the industry needs to establish feedback loops. Community-driven open source implementations demonstrate the feasibility of the defined reference architecture. An MVDS (Minimum Viable Data Space) gives a first impression of how technologies can be plugged together. This is the first step to starting projects for specific use cases and gives feedback to the developer community. ~~The~~ M~~m~~arket needs will drive the interfaces of commercial products and services. The feedback loop between use cases and used data products will improve interoperability.

Distinguish between mandatory (MVD) and optional requirements (discuss essential principles and optional one)

The "Public money, public code" campaign[1] wants legislation to require that publicly funded software developed for the public sector to be made publicly available under a free and open source software license. IDS-G is where the developer community finds the reference implementation of all components - available under free licenses. We recommend hosting all technical developments there and ask to contribute to ~~the~~ further development.

## L Legal

There is a strong connection between political and legal factors. Legislation follows political decisions. Besides knowing the existing ~~legislation~~legislation, the impact of new and planned regulations ~~because of~~based on political developments must be taken into account. Politica~~l~~~~s~~ and social sentiments need to be considered.

Legal fields to bear in mind when sharing data include antitrust/competition, data protection and security, copyright, patents/intellectual property. The European Data Strategy mentioned above~~before comes with~~brings a higher level of regulation~~s~~ to~~n~~ data sharing in the EU,

[1] https://publiccode.eu

Commented [M(9]: I want to discuss software development as journey with required feedback loops

Commented [OG10R9]: Maybe this can also adressed to Chapter 4.1? Here, under "Ethics and Vaues", i think we should focus on the OpenSource, Open, interoperable and standards aspects as prerequisite for community driven technical rules, which creates values, or?

Commented [Sö11]: Input from Task Force Legal Framework.

including the Data Governance Act (DGA),[2] the Proposal for Data Act (DA-E),[3] the Digital Markets Act (DMA),[4] the Digital Services Act (DSA)[5] and the AI Act[6]. If a data space operates globally the legal framework becomes more challenging since each country has itstheir own set of rules and regulations.

Commented [AS(12]: This should go into a footnote since all of them were earlier mentioned in this chapter already.

## E Environmental

Commented [VL13]: From the EU data strategy again: Moreover, making more data available and improving the way in which data is used is essential for tackling societal, climate and environment-related challenges, contributing to healthier, more prosperous and more sustainable societies. It will for example lead to better policies to achieve the objectives of the European Green Deal. At the same time, the current environmental footprint of the ICT sector is estimated to be between 5 to 9% of the world's total electricity use and more than 2% of all emissions, a large part of which is due to data centres, cloud services and connectivity. The EU's digital strategy 'Shaping Europe's digital future' proposes green transformation measures for the ICT sector.

Data usage - harvestingcollecting, processing, or federation - has a huge and growingincreasing impact on our planet. The EU Dedata Sstrategy states that making more data available and improvinggg the way in which data is used is essential for tacklingto address societal, climate and environmental challenges, contributing to a healthier, more prosperous, anddd more sustainable society. It will lead, for example, lead to better policies to achieve the objectives of the European Green Deal. At the same time, the current environmental footprint of the ICT sector is estimated atto be between 5 to 9% of the world's globaltotal electricity consumption and more than 2% of all emissions, a large part of which is due to data centers, cloud services and connectivity. The EU's digital strategy "Shaping Europe's digital future" proposes green transformation measures for the ICT sector.

The choice of implementation design can have a significant impact on the energy consumption of digital tools. We strongly recommend an ongoing assessment of the key components and technology that determine the energy profile of data spaces and services. For distributed ledger technologies, for example, the main factors affecting energy consumption are the ability to control participation and the consensus algorithm. While cryptocurrencies like Bitcoin waste resources, other approaches may be more energy efficient than existing payment systems.

When developing data spaces special attention should be paid consideration should be given to sustainable digital technologies. AI-based services and state-of-the-art data mining technologies can increase resource efficiency, optimize supply chains, improve coordinate tion of sector coupling and thus lower emissions and add value. Avoiding rebound effects through the use ofwith digital technologies is an important goal. Continuous monitoring and sustainable design should ensure that the use of digital technologies has a net positive impact on the climate balancefootprint.

[1] https://www.kocos.com/news/blog/articles/article/2021/10/18/some-facts-about-the-energy-consumption-of-digitalisation

Field Code Changed
Field Code Changed

[2] REGULATION (EU) 2022/868 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act); https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R0868.

[3] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2022%3A68%3AFIN;

[4] https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM%3A2020%3A842%3AFIN;

[5] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020PC0825;

[6] https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A52021PC0206.

Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed

## 2.2 Layers of data space governance

The layers of data space governance (Figure 4) are inspired by the Design Principles for Data Spaces[7] publication. This was developed in the context of the OPEN DEI project funded by EU where data spaces experts teamed up to define cross-sectoral principles for building data spaces.

Commented [AS(14]: Could there be a link to the paper in a footnote? And maybe a short explanation who OPEN DEI is/ was?

Commented [TM15R14]: Added

| Layer | Description |
|---|---|
| **Data space instance governance** | Executes and implements the governance practices and rules of a data space instance. Oversees data space functions and the rules. |
| **Data space ecosystem governance** | Defines the rules for the data space instance. Creates the intra data space trust between collaborating organizations. Complements standardization and regulation focusing on business-driven rules. Defines the inter data space interoperability practices. |
| **Data space domain governance** | Establishes sector-specific data space principles and mechanisms including semantic interoperability and domain-specific regulation. Leaves room for geographical differences while supporting maximum interoperability. |
| **Soft infrastructure governance** | Brings all the generic data space building blocks and concepts together, defines the legal basis and creates the common framework on which all data spaces are built. |

*Figure 4 Four Layers to describe data spaces governance*

## 2.3 Data economy with digital sovereignty

Using IDS based frameworks, services and offerings ~~means~~ guarantees data sovereignty for your business.

There are some common rules and guidelines:

- ~~1. There is a C~~common definition on lifecycle agreements for IDS-based assets, the IDS standards and services. See appendix "Operational Agreements, Life Cycle".

[7] https://design-principles-for-data-spaces.org/

- ~~2. There are some~~ Ggeneral definitions of necessary processes for development, certification, onboarding, operation and usage. See appendix "Operational agreements. Processes".

Typical roles ~~anticipated in~~of an IDS based data space are described in more detail in a following chapter. Some papers will also address the different roles with examples of use cases and business models.

In summary, using IDS with its data sovereignty is a competitive advantage for your own business and quite easy to do, since everything is well prepared. The IDSA website provides all information (https://www.internationaldataspaces.org). A hotline can help with questions (SupportOffice@internationaldatspaces.org).

## 2.4 Governance/legal framework

Relationship of data usage control and other types of control enforcement and legal agreements

~~The~~ EU ~~level~~ policies set the framework for data spaces, but each instance will need additional governance. This rulebook helps you put that governance in place. In this section, we briefly cover the relevant EU regulation for data spaces: DGA, DA, eIDAS2, GDPR, NIS2, others <pls add!>. In chapter 6, we cover the contractual aspects of setting up the governance for a data space instance.

## 2.5 Role models

Roles in this rulebook describe functions, and no status. The model definition of roles should provide clarity about tasks and capabilities and support the understanding of architectures and interfaces. Roles may not always exist in their pure form - mixed forms are often experienced by participants in data spaces - and new or more specific roles will emerge over time. In this section we define the most important and common roles without claiming to be exhaustive. In practice, it has proven useful to first implement the essential roles that are necessary for the data space to function. Three roles should be established first: provider, consumer, and intermediary services.

### 2.5.1 Data consumer (essential)

The term data user means a natural or legal person who has lawful access to certain personal or non-personal data and has the right, including under Regulation (EU) 2016/679 in the case of personal data, to use that data for commercial or non-commercial purposes.

### 2.5.2 Data provider (essential)

The term data holder means a legal person, including public sector bodies and international organizations, or a natural person who is not a data subject with respect to the specific data in question, who has the right to grant access to or to share certain personal data or non-personal data in accordance with applicable Union or national law.

**Commented [AS(16]:** in this rulebook or where?

**Field Code Changed**

**Commented [AT17]:** (Comment by Söntje Hilberg)
I assume the Legal Framework Working Group may contribute to this Section. However, we should ask what should be the expected content. I can imagine that there should be something like:
1) General regulatory framework around data (for example IP, Privacy, Confidentiality etc.)
2) Contractual agreements regarding data usage (e.g. licensing)
What should be the key message here??? There is a chapter regarding usage contracts in the IDS RAM (3.1.4).

**Commented [VL18R17]:** @Söntje: v good point! I suggest here we focus on your point 1, whereas point 2 would be covered in section 6. My key message would be something l

**Commented [M(19]:** •**Patterns**
•**Aggregator** – combining data from multiple sources for computation at one partner (Specialization: Data Trustee)

**Commented [M(20]:** Please review 1.2.2 Roles and Responsibilities in Data Spaces

**Commented [OG21R20]:** there is only Provider, Consumer and Federator as roles. We want actually bring at least one

**Commented [TM22]:** Let's sync with Jogi's work:

**Commented [M(23]:** Data Recipient, Data Processor, Algorithm provider

**Commented [OG24R23]:** Algorithm=Service. Check. Processor=Aggregator, Federator or Service Provider=Check. Recipient=Consumer.Check.

**Commented [AS(25]:** It sounds very 'legal' - could it be phrased more readable or is it necessary to be like this? This could be a footnote to the more readable description?!

**Commented [M(26]:** Data Holder

**Commented [OG27R26]:** in the sense of a Data Provider, or in the sense of a "Data Silo"?

**Commented [P(28]:** Terminology mismatch. You are calling it a Data Provider in the title but in the text it's a data holder.

### 0.1.1 Service Provider (intermediary, operator, aggregator)

**Aggregator** – combining data from multiple sources for computation at one partner (Specialization: data trustee)

**Intermediary** service aims to establish commercial relationships for data sharing between a number of data holders and data users. This is done through technical, legal, and other means; it includes to exercise the rights of data subjects in relation to personal data; it excludes at least the following:

~~(a)~~

- services that obtain data from data holders and aggregate, enrich, or transform the data to add value and then license it to data users, without establishing a commercial relationship between data holders and data users

~~(b)~~

- services that focus on the mediation of copyright-protected content

~~(c)~~

- services exclusively used by one data holder to enable the use of the data held by that data holder, or used by multiple legal people in a closed group, including supplier or customer relationships or contracted collaborations, in particular those who want to ensure the functionalities of objects and devices connected to the IoT (Internet of Things)

~~(d)~~

Formatted: List Paragraph, Indent: Left: 0 cm

- data sharing services offered by public sector bodies that do not establish commercial relationships.

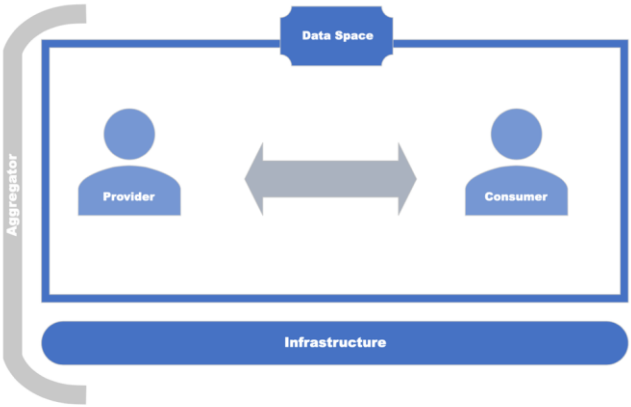



*Figure 5 Roles in a data space*

### ~~2.5.3 Federator~~

~~A federation is a group of participants that works together and collaborates on an equal level. The federation is not owned by anyone; the participants work together based on common rules. One participant of the federation or an external service supplier is appointed to become the so-called federator, who facilitates the coordination of the group and provides the necessary operational federation services. Gaia-X federations based on different industries can consist of a large group of participants.~~

**Commented [P(29]:** This is copied out of Gaia-X for Gaia-X Federations, which assumes that underneath a Dataspace is an extra layer of an infrastructure federation. This should not be in a rule book about dataspaces as it applies only to specific Gaia-X Federations.

**Commented [SS30]:** Peter Koen as maintainer

# 3 Functional requirements for a data space

*This section of the rule book describes the mandatory functional requirements as well as optional elements for building trusted data spaces. It highlights the design decisions necessary to build and operate data spaces in centralized, federated or decentralized architectures and deployment patterns to show how various solutions are enabled by the building blocks of data spaces.*

Enterprises strive to have control over their data. Control is important when managing data internally, but even more in sharing data with others. The core function of a data space is to broker trust between participants and to negotiate available data contracts. They~~to~~ enable control over data sharing and ~~to~~ create value for all involved parties.

A data space is both a multi-organizational agreement and a supporting technical infrastructure ~~that enables~~for data sharing. Participants can have pre-existing levels of trust:

Some may have a prior relationship and trust each other, while others might have no relationship and are untrusted entities. Data spaces even make data sharing between direct competitors possible. -Data space connectors facilitate and orchestrate the sharing of data assets, while enforcing requirements set by the data provider. A connector includes policies, configuration and other metadata artifacts that can run on any cloud infrastructure, on premises or on an edge device.

Data sharing in a data space is not limited to sending data from one participant to another but can be more complex. Fundamentally, all sharing ~~and use~~ of data consists of peer-to-peer interactions. ~~The complex~~All scenarios of multiple actors are built on peer-to-peer data contracts of two participants. A data space adds value beyond individual data transfers by enabling collective data services and applications. These additional capabilities require certain functional requirements to be included in the design of a data space.

Different business, regulatory, legal, or technical requirements ~~will~~ necessitate different architectures and approaches. Some data spaces might require centralized components with centralized control, while others might be designed so their participants have a maximum level of autonomy and maintain agency over how to share their data.

### 3.1.1 Achieving digital sovereignty

Digital sovereignty starts with control over your identity. Identification mechanisms are the basis for finding attributes of a participant in a data space. Identity provides vital information to enable the sharing of data – everyone needs to understand who they are sharing data with. It is the most important function within a data space. It allows the participant to exert control, to choose which data to share with whom, when and under what conditions. This ensures the participant has agency over its ~~own~~ assets ~~and actions~~.

How should the identities for participants be provided? A federated system with a distributed design is a compromise between a centralized and a decentralized design as it enables a higher level of control without relying on a single central point of control. To enable a federated system, services are implemented where multiple participants share the responsibility ~~of providing~~for necessary functionality for all ~~to all participants are implemented~~.

The data space authority (DSA) is responsible for establishing the policies and rules of the data space. This role can be carried out by one entity, but also by multiple or even all participants. In a centralized data space, this could be the operating company. In a federated data space, this function would be performed by the federator(s) agreeing on the rules, while in a fully decentralized data space, various mechanisms are available to the participants. The mechanisms in a decentralized data space enable participants to agree on the set of policies and their enforcement, thus sharing responsibility for ~~the function of~~ the data space authority function.

When evaluating different data space architectures and deployment models, the individual set of rules that serves as the basis is important, regardless of the required services mentioned above. One s~~S~~uch rule set is the book of law for the membership. When a data space ~~is opera~~operate~~s~~ted in a regulated industry, there are laws and regulations ~~imposed on~~for data sharing. In this case, it makes sense to include specific regulations in the data space policy and rule set. This provides clarity when the data space crosses legal jurisdictions or industries.

**Formatted:** Font: 11 pt

**Formatted:** Font: 11 pt

**Formatted:** Font: 11 pt

**Formatted:** Font: 11 pt

**Formatted:** Font: 11 pt

## 3.1 Foundational concepts of a data space

The foundational concepts of a data space:

- Establishing trust
- Data discoverability
- Data contract negotiation
- Data sharing & usage
- Observability
- Vocabularies and semantic models

Additional elements that support these main functions of a data space can include these optional functional areas:

- Application and processing services
- Marketplaces
- Data trustee and escrow services
- Data incubation and service creation

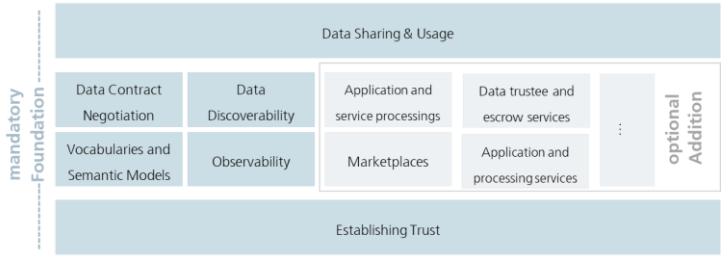



*Figure 6 Foundational Concepts in data spaces*

### 3.1.2 Establishing trust

~~The ability to~~ Eestablishing trust is fundamental to a data space. To create value from data, it needs to interact with other data and then supports decision making. The different entities must trust each other - without trust, data will not be shared. Data spaces can create context-specific trust where trust did not exist before or where it is difficult to establish – for example between competitors.

#### 3.1.2.1 Attributes & self-descriptions

When ~~humans~~ people build trust with~~in~~ each other, they evaluate attributes of the other person: attributes that are immediately verifiable (e.g., a language spoken) or attributes that require an external authority to verify them (e.g., a passport). To build trust, th~~e~~o~~se~~ attributes are matched against (personal) policies. If a sufficient number of policies are met~~satisfied~~, trust is established. Based on the attributes that have been evaluated, different levels of trust can be negotiated.

To create trust in a data space a very similar process is used. It is necessary to evaluate attributes of participants and match them~~ose~~ with the requirements, policies and rules of the data space, the participants, and ~~even~~ individual data contracts.

A data space needs to define policies that specify what ~~level of~~ attributes an applicant must ~~be~~ me~~e~~t ~~for an applicant~~ to become a trusted participant. This is achieved through a data space self-description (DSSD), that allows new members to provide attributes in their participant self-description (PSD) in a format that can be understood by the data space authority (DSA). Therefore, the DSSD must include a reference to a semantic model that describes the acceptable policies, their names, the potential value, and the format in which those values are accepted.

For example, one data space might require self-descriptions to ~~to~~ be expressed as verifiable presentations in a single presentation per attribute, while another data space might require self-descriptions to be expressed as one large file containing all information serialized as JSON-LD for the attributes and ~~applicable~~ corresponding signatures. While participants might manage the values of the PSD through application services which enable complex data management and a permissions system for editing, th~~e~~o~~se~~ services ~~finally need to~~must render the self-descriptions in the desired format that each data space requires at an appropriate service endpoint for that data space.

Trust in a data space needs to be rooted in one or more trust anchors and trust frameworks. These are ~~comparable~~ similar to mechanisms that citizens use in their ~~every~~daily li~~v~~f~~e~~s: The level of trust depends on the authority that issues them, such as a department of traffic issuing drivers licenses or a ministry of internal affairs handing out citizen ID cards. The underlying process ~~consists of~~is verifying a specific attribute.

A trust anchor is an entity that issues certifications about an attribute. The accompanying trust framework is the set of rules imposed by the trust anchor to comply with its policies. Only then is the applicant eligible for its attribute verification. For example, a company ~~based~~ must ~~in a specific country will have to~~ follow the laws of the~~at~~ country it is based in to obtain~~have~~ a valid company registry ID issued by its government.

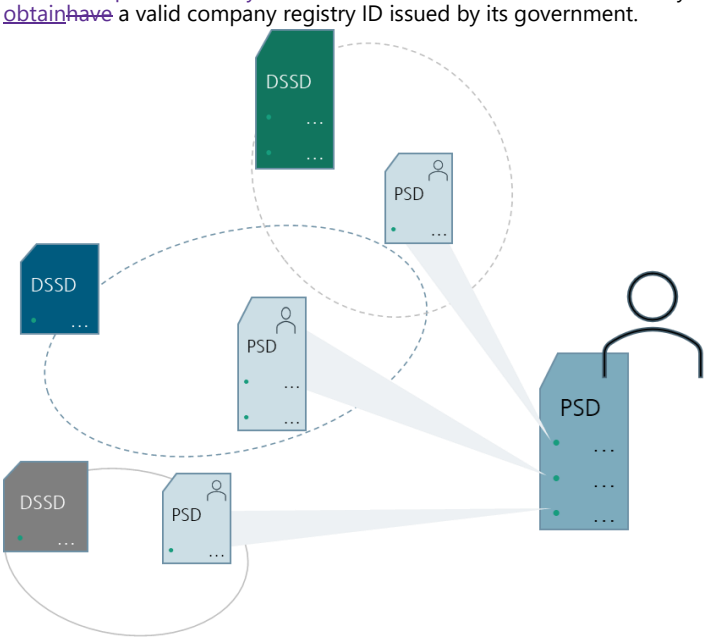



*Figure 7 Self Descriptions in data spaces*

Deciding which trust anchors and trust frameworks, and thus which rules and procedures of issuing and validating attributes are used, is the responsibility of the DSA ~~and also~~and of the participants of the data space. Details can be found in the certification section. For the data space functionality, the concepts of trust anchor and trust framework form the basis for the attribute-based trust mechanism.

In order to~~To make~~ use of the concepts described above, the DSSD needs to contain information about which~~s to what~~ trust anchors and trust frameworks are accepted as roots of trust. ~~Does it act as~~Is it a sovereign entity ~~forming~~ that is the sole root of trust, or is it embedded in a larger ecosystem of external trust anchors and trust frameworks? Based on this~~ information~~, a potential participant can make the decision whether to trust the data space and its members or not.

The DSA is also responsible for issuing membership credentials. It ensures that an appropriate mechanism is provided for identifying and verifying membership~~ is provided~~. In a centralized data space this could be the issuance of a data space specific identity to interact

with other members. In a largemostly decentralized data space architecturearchitecture, it could be the issuance of a tamper-proof credential, such aslike a W3C verifiable credential (VC) which provides proof of the attribute of membership.

The DSA also performfulfils other functional roles not directly related to building trust but necessary for the operation of a data space. These are mainly primarily the mandatory function of regulating the lifecycle of membership (participant discoverability, issuing of membership credentials, verification services for membership proofs), but also many optional services like observability and auditing, brokering and marketplaces, providing vocabularies or other services required by the data space members.

The communities coming together in the data space needs to make decisions for the setup. Whether a centralized DSA is requiredrequired, or a more federated or even fully decentralized model is appropriate must be reasoned over when the data space is founded, as theose architectural choicedecisions are will be very hard to change later. Where on this spectrum of possibilities an optimal design for a data space can be found depends on the context and purpose of the data space.

#### 3.1.2.2 Policies

Policies ensure a trusted data ecosystem within a data space. They are used at multiple levels and at almost any interaction point. The two main policy groups that are central to the functionality of a data space are access policies (which control access to contracts) and contract policies (which control the contract terms and the usage of data). While the use of policies can be expanded by custom design within a data space there are several fundamental policy points that enable the operation and are therefore mandatory essential to be understanood.

It is essential to use policies for attribute-based trust in a data space. Which policies need to be mandatory depends on the design and its requirements. One data space might require policies that reflect the sensitivity of healthcare data in an international setting, while another data space will need to enforce policies for national energy regulation. Therefore, data spaces must define their own policies and clearly communicate them clearly. Participants may always choose additional policies ion their data contracts to further restrict access and use further.

In a centrally managed data space, the DSA might simply define the ontology of policies for the data space. In a decentralized data space, there might be an additional negotiation protocol that enables participants to agree on the policy for their interaction.

Policies generally express three possible restrictionsconstraints: prohibitions, obligations, and permissions. Constraints expressing a rule can be combined into more complex rules, which then form the applicable policy. For example, a group of data space participants may only allow access to their data for participants who belong tojoin the same industry association, allow to process data under the condition only anonymized results are produced, and then

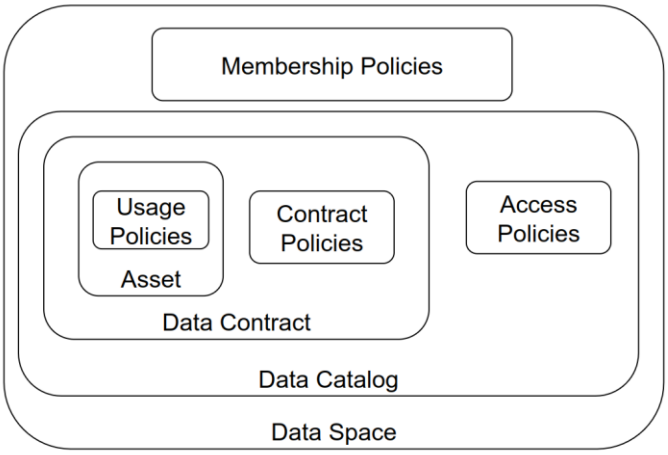



*Figure 8 Different policies in data spaces*

permits to share the results with a third party for processing if they meet a set of ISO standards.

As discussed above, the first line of policy defense is the membership policies (MP) and rules required to join a data space. These policies ensure that only companies with certain attributes they ~~at~~ can verifiably prove ~~them~~, can join. These~~is~~ could be policies that verify the ~~applicants~~applicant's nationality, industry certification, membership in industry associations, but also policies that would require human interactions and complex workflows, such as a valid contract with the DSA that must be negotiated before an applicant can become a participant.

Once an applicant becomes a participant, the next set of policies becomes relevant: access policies (AP). An AP defines which attributes must be available to access data contracts. A participant that does not have access to a specific data contract should also not be able to see the contract offer in the catalog. Optional services, like a marketplace, should adhere to this principle as well and only show items based on ~~the~~ matching ~~of~~ access policies and participant attributes. In a scenario where contract offers should be made visible to~~for~~ everyone, the access policy can also be expressed as an empty policy, not triggering any restrictions. From a functional perspective, an access policy always needs to be present, even if it grants access to everyone. A common scenario is policies that grant access to anyone within the data space but hide the associated item from queries by~~from~~ non-members (in case the catalog endpoint is publicly accessible).

Each participant can define such policies, ~~either~~ whether providing or consuming data. For example, a participant interested in ~~obtaining~~ data could define a policy to see only data with a distinct proof of origin, and participants offering data could restrict access to their data to members of a certain jurisdiction ~~only~~. This is often referred to as provider policy and consumer policy.

When a participant has access to a data contract offer (DCO) the next set of policies comes into play. A DCO can have contract policies (CP) that define what attributes are needed for a

data contract agreement (DCA). CPs ~~are checking~~review attributes that must~~to~~ be provided at the contract negotiation. This could be as simple as ensuring that the participant uses a specific encryption algorithm or software package – both of which could be verified with a technical handshake procedure (e.g., sending a piece of information and requesting the properly encrypted version). A more complex attribute example~~,~~ involving~~cluding~~ human interaction is the association of the data contract with a legal contract between the two parties that typically occurs outside of the data space processes. The negotiation of policies can be ~~anywhere~~ on the spectrum of 100% machine-~~-~~processable and immediate to a human workflow potentially taking a long time.

A contract may also specify policies for the transport mechanism for the data asset transmission: like requiring a protocol, specifying pull or push of data, mandating a data sink in a specific geographic area and other details ~~of the data transfer~~.

CPs may also include usage policies (UP) that take effect after the data is transmitted and control how the data can be used by the receiving party. Depending on the value of the data, use cases, trust levels, contracts in place and many more attributes, there are different possibilities to enforce UPs which come at varying costs.

For data with ~~relatively~~ low importance or data not under a specific legal protection, it might be too expensive to build a system that guarantees control - it may~~can~~ be sufficient to simply monitor ~~the~~ data use and fall back to a legal contract should misuse of the data be detected. Other data might be very sensitive, legally regulated, or costly and ~~therefore~~ require stronger protection and higher technical costs.

When designing a data space and deciding which data to share, it is important to understand the data's classification, and regulatory controls to design not just the right policies but also to mandate the appropriate level of technical components that ensure proper handling of the data.

| **Example** | **Protection Need** | **Explanation** |
|---|---|---|
| Public weather data | low | Some data sets are already publicly available and can be shared without enabling others to derive sensitive data about persons or business secrets. |
| Shipping information | medium | Some data are valuable and at large scale likely to be highly protection worthy as they can give insights into business relations and transactions. |
| Personal health data | high | Personal health data are highly protection worthy due to strong laws and potential danger to the individual in case of data misuse. |
| Machine operations data | high | Industrial data is also usually of high value due to the sensitive business information it represents. |

The atomic expressions of policies can be further broken down into a set of restrictions against which machine-readable attributes can be compared.

#### 3.1.2.3 Attribute based trust

Establishing trust based on attributes is a control mechanism. A participant's level of trust is determined by evaluating participant's attributes, data contract, data asset, and environment attributes. This evaluates the potential risk of sharing data with another participant. This trust level is also based on the participant attributes, the attributes of the data space and the attributes of the data shared in the data space, as well as the applicable trust anchors and trust frameworks. It can express complex rule sets that can evaluate many attributes. There is no limit to the attributes that can be defined and the expression of policy rules to evaluate those attributes.

Depending on the level of risk that can be tolerated for sharing an asset, restrictions need to be put in place. The restrictions are expressed through policies as described above. The proofs of adherence to the policies and rules are expressed through the participant self-description (PSD), as well as additional attributes that might be provided by the participant outside the self-description (e.g., proof that commercial contract for the data exists and that payment for the data has been submitted).

Attributes can be atomic expressions (e.g., the other entity is a participant of a specific industry association) or ~~they can be~~ a set of multiple atomic expressions (e.g., the other entity is under a specific jurisdiction and the ~~target location~~destination for the data transfer in a specific country). Attributes can be compared to static values (e.g., jurisdiction = country) or to one another (e.g., both parties support the same encryption algorithm).

M~~In m~~any situations ~~attributes~~ will ~~be~~ required attributes that are complex and might require complex workflows that~~and~~ can include human intervention. It is not possible to generally answer how to handle extended and complex attributes. This is a question of the design of the data space and its rules.

Attribute based trust provides a dynamic, context- and risk-aware trust model, that enables precise control by including attributes from many different information systems with customized rules. It allows participants flexibility to build and use different implementations based on their requirements.

#### 3.1.2.4 Data space policies and rules

As introduced above, data spaces require membership policies (MP) as first barrier to their data space. There must also be a trust basis to prove compliance with the policy, and an appropriate mechanism to allow each participant to verify that their counterpart is adhering to it. Every data space must define what level of trust is the minimum for members. Each participant can verify other participants membership through a digital signature mechanism provided by the data space or separately verify compliance with data space policies and rules as needed (e.g., if especially sensitive data is shared, all relevant policies and self-descriptions can be evaluated ad hoc to ensure the necessary trust level). Additional trust frameworks (e.g., the Gaia-X trust framework) can be used to provide additional compliance mechanisms. The data space could even be its own trust anchor. The participants decide whether to trust the DSA and its trust anchors.

The first level at which policies take effect in a data space is the membership level. The next level is the catalog: Every participant should only see items in the catalog that match the permission resulting from matching the participant's attributes to the access policies of the catalog. A contract offer should only be visible to those participants who have the right to access it, to minimize unintentional sharing of information. During the negotiation process

for a data contract, the detailed policies of that contract will be applied. Some of those policies may be fully evaluated at that time while others may not be evaluated until later when the data transfer is made or after the data has been received. We refer to these policies as contract policies (CP) and highlight the sub-group of usage policies (UP) because of their importance in data sharing.

**Commented [AS(31):** There is an overlap between this chapter/ paragraph and chapter 3.1.2.4, page 25. Maybe intended? Could be condensed.

It will be impractical for many data spaces to act as the root of trust as they would need to provide the necessary service functions. (e.g., compliance service to verify external attributes). Also, many data spaces will require multiple external roots of trust, whether for regulatory purposes, legal requirements, or simply because of existing trust in established organizations.

A key~~core~~ question of a data space is therefore which roots of trust are considered acceptable and whether any should be rejected. Since this is an attribute of the data space it can be expressed through the data space self-description (DSSD) and its acceptance mandated by~~through~~ the membership policies ~~of the data space , which are~~ encoded in the DSSD.

Another element needs to be part of the DSSD - the mandatory policy information model for the data space. Every data space needs to define the vocabulary to ensure a common understanding of the meaning of the policies~~what policies mean~~. There might be different meanings to the same policy expressions in different data spaces. Therefore, is has to be done individually.

This shows how important the DSSD is for the interaction with the data space functions and to clearly understand the context and ~~the~~ risk factors of the data space. A data space needs to have an identity – not just to be clearly identifiable for the participants and potential members, but also because the identity is the root element to which the DSSD is tied. As mentioned above, the decision on how the functional elements are ~~are~~ implemented and expressed through the functional role of the d~~D~~ata s~~S~~pace a~~A~~uthority is~~are~~ highly dependent on the needs of the data space and is~~are~~ the most important decision to be made when designing a data space.

### 3.1.2.5 Participant information

Information about a participant ~~needs to~~must be discoverable and understandable for other participants - also to enable a clear understanding of the attributes of the participant. Therefore, a participant needs a participant self-description (PSD) that follows a known format and protocol, as well as an ontology that describes the semantics of the attributes.

The format of the PSD can be defined through the DSA and may be a part of the membership policies for the data space. In many cases, the format and ontology of the PSD ~~will~~ also depend on the selected trust anchors and trust framework. For example, ~~for~~ a data space that wants to use Gaia-X as a trust anchor and leverage its trust framework must understand the Gaia-X self-description structure and the meaning of ~~the attributes provided by~~ the Gaia-X self-description attribute definitions. ~~In many cases~~ A~~a~~ data space might require multiple self-description ontologies (e.g., one~~a~~ trust anchor specific ~~one~~ and one ~~an~~ industry specific ~~one~~) which can lead to ambiguity or conflict of definitions, which ~~would~~ have to be resolved by the DSA.

The technical representation and communication of the PSD may vary from one data space to another and will be influenced or mandated by the trust anchor(s). One trust anchor and its trust framework might require attributes to be presented as verifiable presentations when queried, while another might require the possibility to request a set of attributes serialized in

a specific resource description format, and a third one might require that all attributes be made discoverable in a database that's available to all members for query at any time.

Entities that are participating in multiple data spaces at the same time ~~need to find a way~~must ~~to~~ manage their self-description attributes in a way that reliably keeps attributes up to date, but also filters which ones should be available in which data space and serialized in which format. For larger enterprises with complex roles and responsibilities related to the information contained in the attributes, this might include approval processes and audit functions to track value changes to sensitive attributes exposed by the self-descriptions.

Information exposed through participant self-descriptions (PSD) is used in many policy evaluations throughout the data space. A non-exhaustive list of examples is:

- Information for the registration process to evaluate whether an applicant can become a participant.
- Matching participant attributes to access catalog policies to only show items this participant is permitted to see.
- Automated matching of attributes to policy requirements in the contract negotiation process.

Commented [AS(32]: This seems a repeat of earlier paragraph. Can it be deleted here?

Self-descriptions can also be used to convey purely technical information about a participant. For example, at what address can another participant communicate with its catalog or connector with this participant, what encryption techniques are supported. Whether this information is stored and distributed in the same way as the PSD is a question of the data space design. A data space that is using centralized components for all mandatory functions will not require a per participant discovery mechanism, while a more decentralized design will require some discovery functions that can be implemented through the same mechanism as the PSD or possibly through separate protocols.

### 3.1.3 Data space participation

Participation in a data space is based on fulfilling all the policies, rules and procedures that are mandatory for ~~data space~~ membership. In its simplest form, these~~is~~ ~~can be~~may just be technical ~~policies~~ or automatically verifiable policies ~~that are automatically verifiable~~. In more advanced cases, these can be more complex policies and rules that potentially require l~~o~~ngthy ~~running~~ workflows with human interaction to verify eligibility to ~~for participation~~join ~~in~~ a data space (e.g., a signed legal contract with a central operating company, membership in industry associations).

The procedure to join a space will likely include the following steps for the applicant~~When a candidate wishes to join a data space the procedure to do so will most likely be something along the lines of the following steps~~ (details can vary due to the design and purpose of the data space):

- Candidate~~s~~ discover~~s~~ the data space and the corresponding DSSD
  This can be achieved ~~by numerous channels. Possibilities include~~ through human

Formatted: Bulleted + Level: 1 + Aligned at: 0,63 cm + Indent at: 1,27 cm

interaction, a website of the data space, finding the DID[8] of the data space in some ~~form of~~ registry or through automated discovery protocol of existing participants among other things.~~.~~

1. Candidate reads the DSSD and receives information about the policies and rules of the data space, as well as technical configuration information for endpoints and protocols.
2. Candidate evaluates the policies and rules ~~of the data space~~ and prepares additional information needed for the requirements when applying for membership in the data space.
3. ~~Once~~ When all information ~~has been collected~~ and necessary proofs are collected~~gathered~~ the candidate applies for membership through the registry function of the DSA. The ~~exact~~ technical implementation of the data space registry might vary based on the requirements.
4. The DSA requests proofs for all policies ~~that are required to join~~. This might include VCs and proof of technical capabilities, but also workflows including human interaction (e.g., signing a membership contract).
5. Once all policies have been satisfactorily processed the DSA issues a VC/ proof of membership and sends it to the candidate, moving them from applicant to participant.
6. The new participant sets up all the necessary technical components for participation in the data space.

[8] https://www.w3.org/TR/did-core/

7. The application process is complete, the participant can start interacting with other participants (sharing data, browsing the catalog(s) for data of others, negotiating data contracts).

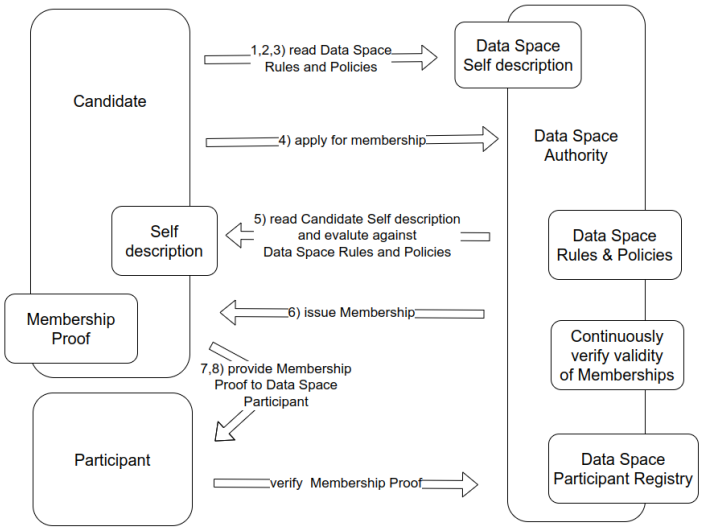



*Figure 9 Onboarding in data spaces*

### 3.1.4 Creating a data space

~~Now that we have~~After discuss~~ed~~ing how to join a data space the question is: How do you create a data space? The answer ~~is, as so many times in this document, it~~ depends again on the purpose of your data space and the needs of its participants. R~~But r~~egardless of whether the data space is organized in a centralized, decentralized, federated or hybrid manner, common denominators and basic functionalities can be found.

A data space establishes trust within a community to~~that wants or needs to~~ share data with each other. The definition of community can be very broad. It might be a tight knit, small community of one company and its suppliers, ~~but it also might be~~or a ~~rather~~ large community with many participants. Some data spaces ~~will be~~are created for a ~~very~~ narrow use case and purpose others for many use cases that are relevant for ~~the same~~a group of participants.

Many decisions need to be made when designing the data space, here some of the more common ones ~~are on the following spectrum~~:

- Is the membership closed to a small, known ~~known~~ group or open to a larger ~~wide~~ range of participants?

**Formatted:** Bulleted + Level: 1 + Aligned at: 0,63 cm + Indent at: 1,27 cm

- Do you want ~~to~~ a central party with additional privileges (e.g., exclusion of participants for bad behavior) or is the independence of the participants and their autonomy the most important design factor?
- What level of technical maturity is~~will be~~ expected from the participants?
- What type of data is~~will be~~ shared and for what purpose?

Answering these questions ~~will~~ helps you make the ~~necessary~~ design choices between architectures and deployment patterns of data spaces.

Once all design decisions ~~have been~~are made, the functional elements ~~need to be~~are planned:

- Rules: What behavior and skills (technical and organizational) are required?
- Policies: the participation rules expressed and verified in policies
- Membership certification: What mechanism is used to verify a membership?
- Participant registry: Where can participants see who is participating?
- Identity system: centralized or decentralized identities - control over participants
- Catalog(s): one central, multiple federated or individual decentralized catalogs?

Working through the above list of mandatory functional elements will ~~solidify~~ clarify the architecture pattern ~~chosen~~ for the data space, which will also mandate a specific design of the d~~D~~ata s~~S~~pace a~~A~~uthority. Now ~~that all decisions have been made,~~ the DSA needs to be implemented to create the data space:

1. Create an identity for the data space
2. Provide a self-description
    - Membership policies
    - Trust anchors and trust frameworks
    - Attributes that will help participants decide which level of trust to apply for
- use of the technical components as required according to the design
- Participant registry
    - Registration service

**Formatted:** Bulleted + Level: 1 + Aligned at: 0,63 cm + Indent at: 1,27 cm

**Formatted:** Bulleted + Level: 1 + Aligned at: 0,63 cm + Indent at: 1,27 cm

- Provide the workflow to apply for membership
- Validate whether applicants comply with membership requirements
- Issue membership credentials
- Revoke membership credentials

3. Provide a discovery mechanism for the data space (website, contact form, etc.)

Once the DSA is instantiated, organizations can apply for membership. After a participant joins, there are two main activities that all participants are interested in: discovering data shared by others and sharing their own data in a controlled manner to ensure autonomy and agency over the data. This is the core functionality that any data space ~~must~~ provides. Additional functions and services such as marketplaces, data escrow services, processing services and applications might be provided as optional elements.

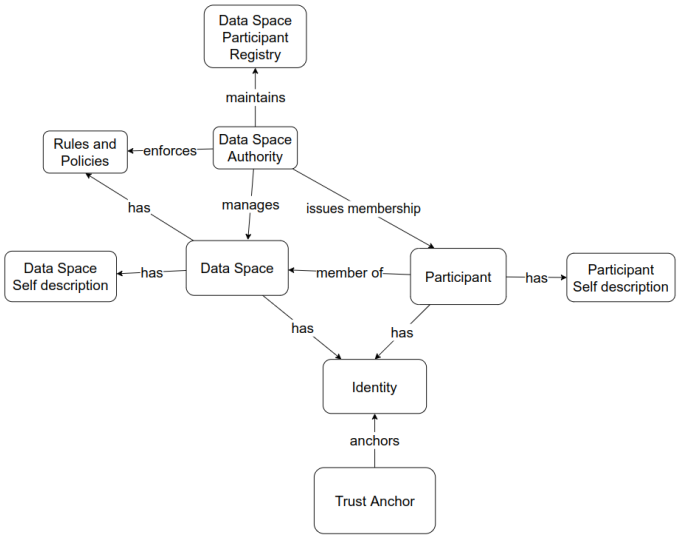



*Figure 10 Relationships and concepts*

### 3.1.5 Data discovery

Regardless of the architectural design of the data space, the most used function is the discovery of data shared by other participants. While the detailed technical mechanisms vary for each implementation and design, there are several common functional elements that are mandatory for all implementations.

### 3.1.6 Catalog(s)

Sharing data among participants requires the provision of metadata – regardless of the design of the data space (centralized, federated, or decentralized) and whether~~of~~ the data is~~being~~ open or protected. Information about the ~~existence of~~ data needs to be published with an agreed-upon vocabulary for querying ~~items~~ and with controls that regulate access to the catalog items.

T~~While t~~wo participants can share data directly ~~without the need for a catalog~~ communicating off- or online without the need for a catalog~~(e.g., by agreeing on the data sharing and the address of the data in an offline channel)~~. But for more participants ~~having~~ a catalog function greatly increases the discoverability of data assets and services. If there is more than one catalog due to a federated or decentralized design, the catalog ~~needs to~~must allow federated searches of~~for~~ data assets in~~across~~ catalogs at~~in~~ multiple sites.

Catalogs don't provide the data asset itself, but they provide ~~are~~ data contract offers (more on this in the section on data sharing below).

When choosing a target architecture for a data space, the design of the catalog function can fall somewhere along the spectrum between a central catalog, multiple federated catalogs,

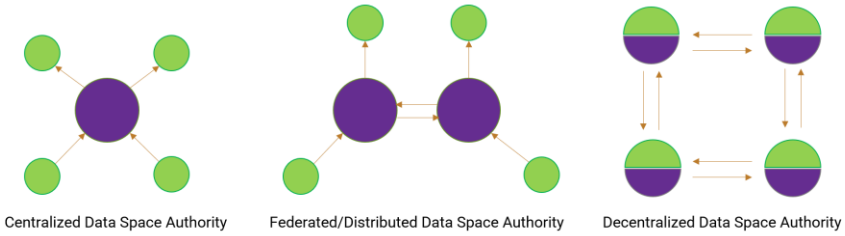



*Figure 11 Variants for data space authorities*

and many decentralized catalogs. Each has its own advantages and disadvantages. Compare the three main types of catalogs, depending on the implementation design of the DSA, to evaluate their capabilities~~.~~

~~When choosing a target architecture for a data space the design of the catalog function can be anywhere on the spectrum between a central catalog, multiple federated catalogs and many decentralized catalogs. Each has its own advantages and disadvantages. Compare the three main types of catalogs, depending on the implementation design of the DSA, to evaluate their capabilities~~:

| Catalog architecture | Advantages | Disadvantages |
|---|---|---|
| **Centralized catalog** | • No deployment by individual participants | • A central gatekeeper can arbitrarily exclude |

| | | |
|---|---|---|
| | • Central control – a gatekeeper can regulate which entries are permissible and which are not<br>• Easy discovery as only one catalog needs to be queried | participants and their data from the catalog<br>• Single point of failure<br>• Potential performance bottle neck<br>• Security issues will affect all members at once |
| **Federated catalog** | • Deployment by a limited number of participants, while most participants don't need to deploy any catalog components<br>• Federated control – voting mechanisms for content control can be implemented | • Additional replication mechanisms are needed<br>• A small group of operators of federated catalog nodes can control participation in the data space |
| **Decentralized catalog** | • Every participant can autonomously decide which catalog items they share with whom<br>• No interference in the interaction between two participants through a 3rd party<br>• Data sSpace as a whole is more resilient towards cyberattacks even though individual members can experience outages<br>• Easier to scale | • Every participant needs to run a catalog component<br>• A list of available catalogs needs to be either centrally provided through the DSA or discoverable through a peer-to-peer protocol<br>• Participants need to ~~crawl~~ search each other's catalogs to see which items are available |

#### 3.1.6.1 Access policies

A best practice of access security is for an IT system to show users only what they need to know - to minimize the potential attack surface. The same is true for data contract offers (DCO) in a data space: Participants should only see the DCOs for which they are authorized to request a contract negotiation. This does not imply that the participant already has authorization for the data but only ~~means~~ that a participant is allowed to see that the data exists. The permission to access is part of the data contract negotiation. Any catalog ~~needs to~~must implement attribute-based access control (ABAC) through access policies.

The most common access filter is that a participant ~~must~~ proves membership to see which assets are in a data space. Filters can also be applied that make data assets accessible only to specific participant groups. For example, a participant who has a VC as a data space member, but also has an additional VC which attests that the participant is ~~acting as~~ an auditor, could provide this participant access to audit log files or streams which are being shared as DCOs, but should not be visible to participants without the special auditor credentials.

In case a participant wants to make a DCO visible to other entities that are not participating in the data space and are merely using the technical mechanisms of the data space or have been directly informed about the existence of those DCOs, they could have an access policy which is simply a no-op, or allow-all policy.

Access policies can also be used as filters to control visibility/access to DCOs. For example, time-based policies can be used to control when DCOs can be negotiated, location-based policies can limit the audience to participants from a specific geographic region.

### 3.1.7 Data sharing

Once a participant has joined a data space and discovered available data contract offers, the ~~main~~ mechanism of data sharing is initiated. D~~The d~~ata sharing ~~process~~ is the core activity to enable further data processing and value generation by using the data.

Data sharing is a very broad term in this context. It ~~can be~~ranges ~~anything~~ from a one-time transfer of a file, access to an API, registering for an eventing service, subscribing to a data stream, also including data sharing methods where the data remains ~~at rest~~ at the source and algorithms and processing code are copied to the data location for in-place processing. Data Sharing does not ~~necessarily~~ require a physical move of the data asset, although this will be frequently the case.

However, before data can be shared, a data contract offer needs to be negotiated to reach a data contract agreement (DCA) which specifies all policies and details of the data sharing process.

#### 3.1.7.1 Contract negotiation

A contract negotiation (CN) serves the purpose of reaching an agreement to share a data asset between two participants of the data space. During the CN policies of the DCO are evaluated against the attributes of the requesting participant, and VCs are verified with their issuers. Note that while any trust anchor is an issuer of VCs that can be used to evaluate policies, there might ~~also~~ be additional external issuers that need to be validated (e.g., government agencies, regulators, industry associations)

It is important to note that the CN does not automatically lead to an immediate data or algorithm transfer. The result of a CN is a data contract agreement, which then can be executed at a later point in time.

Imagine a scenario where ~~in a large enterprise~~ multiple roles are involved in the process of data sharing in a large enterprise. The person negotiating the DCA might not be the same one who ~~then~~ is responsible for sharing the data. Or there might be data assets that can't be immediately shared ~~when~~ after the agreement is reached (e.g., an event notification that can only be consumed until~~once~~ the event in questions has occurred).

**Data sharing execution**

When it is time to share the data, it might be necessary to re-validate the policies of the data contract agreement as significant time might have passed since the contract negotiation. The decision whether to revisit all policies might depend on each party's business rules. If data needs to be highly protected or requires specific regulatory processes for handling it, it is advisable to conduct an additional review.

To exercise ~~the data of~~ a data contract agreement (which could also be code to process data), data needs to be moved from one participant to another. This can be done either by a push model in which the participant with the data asset pushes the data to the other participant or by a pull model, in which the data asset is made available to the consuming participant via a link.

The data transfer technology depends on the type of data asset, trust level, availability of technical protocols, infrastructure environment, and other factors. All data transfer technologies must be able to be orchestrated. Orchestration at this level means having technical control over the data sharing process, allowing the connector to start and stop the transfer, as well as having the necessary technical capabilities to monitor the progress of the transfer and to receive information about compliance with usage policies.

The transfer itself needs to ensure security, performance, and manageability. For example, a data stream can be provided from multiple data centers to enable a highly available data sharing architecture.

When data is not moved but a "code to data" approach is selected, the push and pull behavior is reversed: The consumer participant provides a data asset containing code (source code, compiled library, signed container) to the participant providing the data. This can be implemented like any other data asset transfer with a push or pull mechanism.

Data sharing ~~needs to~~must accommodate a wide range of scenarios. From a simple file transfer between two storage providers, to API access for streaming or eventing, to quite complex implementations with secure execution environments through confidential compute enclaves, environment attestations, signed code, custom encryption algorithms, and more. Which solution is right depends on the protection ~~-~~needs of the data and the trust level between the participants.

The transfer technology can be specified as a policy i~~o~~n the data contract agreement, or it can be implicitly inferred by the type of data asset being shared. A participant who wants to ensure that data never leaves an environment where full control over its usage is guaranteed

can enforce the selection of the transfer technology and storage and processing infrastructure by setting policies in the contract and monitoring compliance.

### 3.1.8 Observability

In data spaces with highly regulated data, it is necessary to make the data sharing process observable. This can be done for legal reasons to prove that data has been processed only by authorized entities, or for business reasons to provide a marketplace and billing function through a trusted third party.

Depending on the architecture of the data space, multiple solutions are possible. For a centralized architecture a central observer (sometimes called clearing house, auditor or monitoring agent) can be implemented. But this design has ~~some~~ two shortcomings when implementing large-scale data spaces:

- It presents~~is~~ an additional vulnerability that could affect the sharing of mission critical data. And a

- ~~A~~ central observer has data on all DCAs which represents potentially valuable knowledge about the participants. This can be exploited for financial gain, making it a target for bad actors.

Formatted: English (United States)

Formatted: Normal, No bullets or numbering

Formatted: English (United States)

To address these risks, having at least a federated model of observers is recommended to distribute the information, load, and potential for ~~failure~~error ~~is recommended~~. To go a step further, a decentralized architecture can minimize the risks associated with a centralized or federated observer model.

In a decentralized observer architecture, every participant keeps the information about the agreed DCAs and their execution in their own environment. Meaning that there are at least two copies of corresponding logging information in the data space. The two copies can always be identified through a correlation ID linking them ~~together~~. The observer then matches the corresponding logging information and reports any irregularities to the parties participating in the DCA (or to the respective regulator if required).

A third party participant in the data space can have an additional VC which qualifies them as a trusted observer, such as an industry auditor, rooted in a governmental trust anchor for auditors.

To audit the contracts of a participant, the auditor would simply request the log data which could then be published as data contract offers with an access policy which restricts access to the auditor. To verify the validity of those log entries, digital signing mechanism can be used or the corresponding log data from other participants can be requested (and again published as data contract offers). This would limi~~restric~~t access to sensitive observation data to observers that are participants of the data space, have special credentials which qualify them as trusted auditors and~~, due to the contracts on the collected log data,~~ are bound to the policies of those contracts due to the contracts on the collected log data. Observer actions are automatically logged by the system and can be tracked and monitored. This would enable a trust relationship in which auditors can be audited by participants.

To simplify the observability of a data space, the DSA can mandate that participants make their audit data available as events or streams per default. Then trusted auditors would not need to request publication but could simply negotiate the relevant contracts, which are only accessible to participants with valid auditing and monitoring credentials.

Following the same pattern, additional optional functional roles can be implemented. For example, a payment clearance service, notary services, regulatory reporting, and the like.

### 3.1.9 Vocabulary

Vocabularies are used to ensure that everyone means the same thing when using a specific term. There are multiple vocabularies that are needed in a data space, but two are particularly important:

- Semantic models for policies
- Semantic models of the shared data assets

So far, this document mostly described how a data space works, what contracts are, what types of policies exist, and how to negotiate a contract. The vocabularies describe the content of these elements.

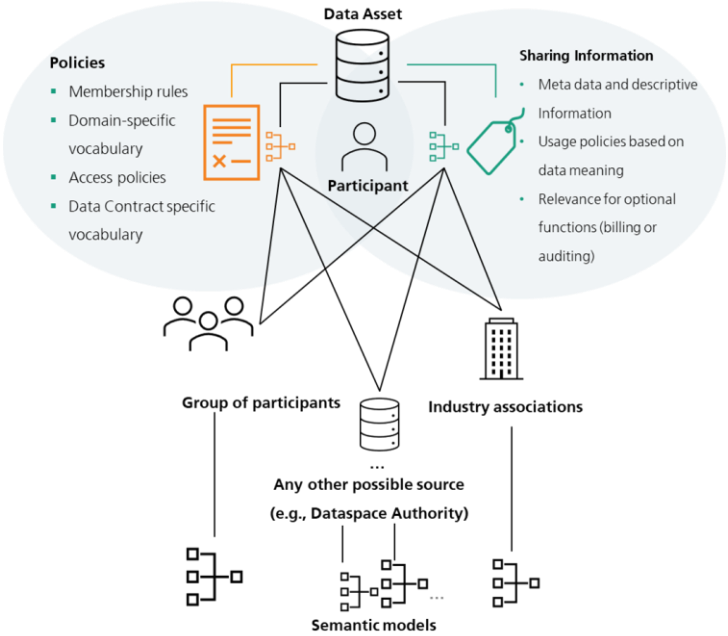
The first category is the vocabulary of policies, which can exist on multiple levels:

- Semantic model for policies for membership rules
  E.gFor example, if a data space wants to restrict membership to companies with a HQ in certain countries. It It must be clear what the policy is called and what values are allowed.
- Policies that each member of the data space must understand to interact with other participants. For example, policies that specify which industry vocabularies must be understood, and access policies.
- A participant can publish additional information on semantic models relevant for the interaction with this participant. This could be special access policies under which this participant publishes additional contracts. It could beE.g., an access policy that specifies access for direct suppliers of this participant.
- Data contract
- Semantic model which needs to be understood for a specific contract (e.g., special usage policy for a single contract)

Formatted: Bulleted + Level: 1 + Aligned at: 0,63 cm + Indent at: 1,27 cm

The vocabularies for each level can be easily referenced by the metadata publishing mechanism at the respective level. A data space can reference the required policy vocabulary through its self-description. A participant can also leverage its self-description to publish additional vocabulary requirements. And at the data contract level, this information can be easily stored in the metadata associated with the contract at the catalog level.

For mandatory vocabularies a policy referencing them can be easily established if such a policy model has been agreed upon.




Semantic models for data assets work on the same principle with the main difference that they do not describe functionality of the data space itself, but the meaning of the data being shared. If this data needs to be understood to properly handle usage policies (e.g., if usage policies are based on the meaning of data) it becomes an essential part to be considered in the design of the data space. Semantic data models might also be relevant for optional functions such as billing and auditing.

How best to manage the publication of vocabularies depends ~~highly~~ on the design of the data space and its requirements. There can be central servers hosting the semantic models, public semantic models from industry associations that can be referenced externally, a group of participants responsible for publishing and synchronizing common semantic models, or semantic models that each participant receives when joining the data space and which can be continuously updated through various synchronization mechanisms.

### 3.1.10 Optional functions

In addition to the functional elements of a data space, many optional roles and components exist. The entities providing these functions must join the data space like any other participant and fulfill all requirements, policies and procedures enforced by the DSA to establish ~~create~~ trust.

*Figure 12 Vocabularies and their relationship to data assets*

Depending on the services provided, these additional elements ~~might have the need~~may need to issue additional credentials, introduce additional trust anchors, or require specific data contracts. There is a ~~great~~ wide variety of optional roles and services. Some especially useful ones are described here.

#### 3.1.10.1 Marketplaces

Commented [VL33]: Data exchange platform?

Commented [VL34R33]: "Marketplace" confuses the technical means of data exchange and the business of value exchange (incl. the euros)

Data sharing always takes place peer-to-peer in a data space with data discovery being provided via catalogs. This basic functionality does not cover any form of business model. Since~~As~~ many dataspaces require not only ~~will require not just~~searching for ~~the discovery of~~ available data but also platforms for trading, buying, and selling data, it is expected that many different models of data marketplaces will emerge within data spaces.

Again, these can be centralized marketplaces, federated marketplaces, or individual decentralized business platforms. S~~It is s~~imilar to how resources can be bought and sold on exchanges, ~~similar~~ function~~sality~~ can be created for data contracts. A marketplace can also provide a catalog that enables data discovery as well as a business platform to buy and sell data. Or it simply may act as a broker facilitating the negotiation of data contracts for a fee.

#### 3.1.10.2 Processing services

A data space can have participants that do not offer their ~~own~~ data and are not the ~~final~~ end users ~~consumers~~ of data. At its most basic level, these can be participants that are offering algorithms and code for processing data as a data contract to deliver code libraries, signed containers, or entire virtual machines to other participants. For very computation intensive or special hardware requiring workloads th~~e~~~~o~~se participants might offer their own infrastructure as part of the contract and use policies to control the use of their resources.

Many ~~models of~~ data spaces can be built on top of the peer-to-peer model, ~~for example,~~ such as a data supply chain w~~here~~~~ith~~ data assets ~~being~~ pass~~ed~~ through multiple processors before ~~arriving at the final cons~~reaching the end user~~umer~~. The implementation and capability of these services ~~is~~ again depend~~sent~~ on the architecture, policies, and rules of the data space.

#### 3.1.10.3 Data escrow, data trustee

For many applications, data assets and algorithms from multiple sources need to be combined to generate value. This will lead to trusted service providers collecting all necessary data, perform the calculations, and then distribute the results - while adhering to all contract policies and guaranteeing the execution of usage policies such as the enforcement of deletion rules. The business model for these participants will be only ~~be~~ to provide trusted services and not t~~o~~~~he~~ use ~~of~~ the data.

Plenty of possible models are conceivable, from centralized, federated to decentralized offerings with different technical capabilities, trust levels and costs. C~~Even c~~lassi~~cal~~ data aggregation platforms such as data lakes can also be a possible implementation and benefit from the trust which a data space provides.

## 3.2 Technical components of a data space

### 3.1.11 Data space authority services

Several services are required that represent the functional role of the data space authority (DSA) to enable the management functions of a data space. These services may be designed as centralized, federated (distributed) or decentralized services (See below for more information on the differences between these solution designs). Depending on which design is chosen, these services can be implemented with varying component designs that best support the needs of the data space.

Regardless of the technical implementation and the specific architecture model, the following components are required:~~Several services representing the functional role of the data space authority (DSA) are required to enable the management functions of a data space. These services may be designed as centralized, federated (distributed) or decentralized services. (See below for more information on the differences between these solution designs). Depending on which design is chosen, these services can be implemented with varying component designs that best support the needs of the data space.~~

~~Regardless of the technical implementation and the specific architecture model, the following components are required:~~

- Registration: A service providing the requirements of the data space to apply for membership (includes the validation of attributes and their values of the participant self-description and checking their applicability against membership policies). This service can be machine based but can also include human workflows.

- Membership credentials: a membership issuance and verification service can be used to manage membership credentials. Also responsible for revocation of credentials.

- Participant directory: Enables the discovery of other participants in the data space.

**Formatted:** Bulleted + Level: 1 + Aligned at: 0,63 cm + Indent at: 1,27 cm

### 3.1.12 Identity

The design of the identity provider is the first decision for the design of the data space. If a central identity provider is chosen to~~, which~~ manage~~s~~ the identities for all participants, ~~then~~ ever~~an~~y other service ~~will~~ depends on this central verification, and decentralized designs are~~will~~ no longer ~~be~~ fully ~~implementable~~feasible.

~~Choosing~~ W~~w~~hich mechanism to~~is~~ use~~d~~ to identify participants is the most fundamental design decision. It impacts policies on autonomy and sovereignty as well as technical solution architectures for other components of a data space.

| Identity System | Advantages | Disadvantages |
|---|---|---|
| Centralized identity provider | • Simple management for DSA | • Low autonomy and sovereignty of participants |

| | | |
|---|---|---|
| | • High degree of control for DSA<br>• Traditional, well-known technology stack | • Single point of failure<br>• Single point of attack<br>• Harder to manage for participants |
| Decentralized identities | • Full autonomy and sovereignty for participants<br>• Low resourcing need for DSA<br>• Easy to manage for participants<br>• Resilient<br>• Harder to attack | • Complexity: DSA management requires decentralized protocols<br>• Lower degree of control for DSA<br>• New and partially unfamiliar technology stack |

### 3.1.13 Catalog

The catalog component supports the search for available data contracts. Information about data contracts can ~~potentially~~ be exchanged between participants without the use of a catalog by ~~directly~~ sending the offer directly via a separate channel (e-~~-~~mail, notification). A catalog will be a common component to implement data discoverability. It can be implemented as a managed service by one or more selected participants, hosted by the data space authority, or operated in a fully decentralized fashion by every participant that offers data contracts (see the visual representation of various implementation designs of the DSA above). The type of catalog architecture used depends on the design of the data space as well as the needs and capabilities of the participants.

Hybrid catalog models combining central and distributed catalogs with individual decentralized catalogs are possible, but must be carefully designed to avoid unnecessarily increasing the complexity of participating in the data space.

#### 3.1.13.1 Attributes & self-description

Attributes and self-description should always be available as verified presentations. The exact serialization format and service endpoints depend on the implementation of the data space and the trust anchors in use.

### 3.1.14 Connector

The connector forms the gateway for a participant to a data space. It provides the necessary API endpoints for other participants to negotiate data contracts and request the execution of a data contract. The connector acts as an agent of the participant to the data space.

Which solution components are provided by the connector beyond the contract negotiation and execution depends on the implementation design of the data space.

### 3.1.15 Observer

As described above, there is no specific technical component for an observer as this is a role within the data space and not a component.

### 3.1.16 Vocabulary

The semantic model for the policies and self-descriptions required to join the data space is provided by the DSA. It may also provide semantic models that need to be understood throughout the data space and might be mandatory for the publication and use of specific data contracts.

The DSA must decide how semantic models are provided, whether by reference to a known, standardized schema externally or through a vocabulary service provided by the DSA or specific participants.

Individual participants may~~ight~~ provide additional vocabulary services to enable the discovery of semantic models needed to successfully share data with that participant. These could be additional semantic policies or semantic models that describe the shared data model. For example, the semantic model of the shared data must be understood by the consumer to properly manage consent for GDPR.

As mentioned before, the importance of the implementation design of the DSA and the components of a data space cannot be emphasized enough. The implications for autonomy, sovereignty, reliability, security, and many other factors are far reaching, so the decision on the design needs to be made with utmost care.

### 3.1.17 "Central," or "f~~F~~ederated/d~~D~~istributed," or "d~~D~~ecentralized"

**Centralized data space authority**

In a centralized DSA design, the entity runs all services to operate the data space. These include services to identify participants, onboard new participants, manage memberships, provide semantic models, discover data and optional services like marketplaces and audit~~sing~~.

While this model is popular due to the familiarity with centralized models through existing aggregator platforms, it limits the autonomy and sovereignty of participants. If a centralized identity provider is used, the entity that controls the identity provider also controls membership and access to resources. This entity could make arbitrary decisions on inclusion or exclusion without regard to the policies of the data space. Worst case, such a central

identity service could interfere with the data sharing between two participants, with serious consequences beyond the data space.

A central catalog has advantages for data discovery as it provides a~~one~~ known location to discover available data and queries only need to be made ~~against~~ at one end-point and data contract offers are returned from multiple participants. But it poses the risk that the entity controlling the catalog ~~can~~ also controls its content and make arbitrary decisions which items are available to whom.

Centralized services also create a single point of failure. Outage could result in the entire data space becoming unavailable or inoperable. This could cause a significant business risk for participants.

If the data shared is valuable data that should be highly protected, it could ~~might~~ attract bad actors trying to gain access, manipulate it or simply disrupt operations to harm their targets. When a lot of value is aggregated into a centralized component, it could become the target ~~for bad actors~~. An infiltrated central identity provider or catalog could create~~create~~ more damage than if a~~n~~ single participant is attacked.

With careful planning and the right choices when implementing a centralized data space, many of the issues that can prevent participant autonomy can be avoided or softened. But vital functional resources of the data space do not allow for full autonomy of participants in this design solution. However, depending on the purpose and goals of the data space this may~~ight~~ not be a problem.

**Federated / distributed data space authority**

The federated or distributed model retains some degree of centralized control but improves on the technical and security challenges. In this model, functional roles are distributed to a few federated nodes. Instead of just one entity providing a service, multiple entities share responsibility for providing this service through individual nodes that are synchronized. This requires some additional technical investment as nodes need to be synchronized, transactions handled, and queries performed across multiple services.

While this model strongly improves resilience and availability, it also increases complexity. Some functional roles are more complex to implement in a distributed environment (e.g., identity) than others (e.g., catalog). However, it offers interesting variations on the centralized design by allowing more sophisticated designs. For example, a federated catalog could be implemented so that different sub-catalogs are available on different nodes, instead of synchronizing all entries everywhere, increasing performance and availability of the system.

If the goal of the data space is to maximize participant sovereignty and autonomy, the distributed model does not provide significant improvements in comparison to the centralized design because a small group of entities would have most control over the data space and the participants would be almost as dependent on these entities as in a centralized data space.

Nevertheless, a federated model can be the optimal solution to implement data spaces based on closed group consortia with clear consortia leaders. There may be reasons beyond the technical design, such as contracts and legal regulations that necessitate implementing a data space as a federated or partially federated model.

When talking about distributed data spaces there is a distinction between "*Federation service*" and *"Federated service"*.

- Federation service
  ~~It~~ supports the federation functionality of a data space and serves a functional role such as identity or catalog.

- Federated service
  ~~It~~ describes the implementation of any service as a distributed service in a data space, including but not limited to any of the federation services.

Formatted: Bulleted + Level: 1 + Aligned at: 0,63 cm + Indent at: 1,27 cm

To maximize the sovereignty and autonomy of participants in a data space, every participant must be freeable ~~freely~~ without being improperly impeded by anybody. A participant must follow the rules and adhere to policies, but a sovereign participant needs to be immune from undue or random interference. Improper interference can include refusal to put a participant's data assets in the catalog despite meeting all requirements or deactivating the participants identity and thus potentially disrupting the participant's business. This may not be malicious interference; errors can ~~happen~~happen, and the software could be unstable. A fully sovereign participant must be able to interact with other participants without depending on a third party once it is proven that the participant is following all rules.

**Decentralized data space authority**

Using a decentralized design enables the highest level of autonomy and sovereignty. The core element enabling a participant to act autonomously is the identity system. By using a decentralized identity system each participant is responsible to maintain identity information that can be verified by other participants or the DSA, rather than relying on a centralized identity provider.

Once decentralized identities are established, all other functional services can also to be decentralized, minimizing or even eliminating ~~the~~ barriers to participant sovereignty.

It should be noted that in a decentralized data space a lot of the responsibility for operating essential functional roles shifts from the DSA to the participants. For example, in a centralized model, the DSA is expected to operate the catalog of available data assets, while in a decentralized model, each participant is responsible for publishing its available data directly and in turn, each participant needs to ask all other participants ~~what~~ about their available assets ~~are~~.

Another advantage of a decentralized system is that it is usually more resilient to errors or bad actors, since problems in individual nodes do not automatically affect all participants of the data space. Finally, a decentralized system does not require an ever-~~growing~~ increasing number of centralized services. Each node is self-contained and provides all the endpoints necessary to interact with it. A data space can grow and scale much more efficiently than a centralized design, where the resources to provide central services must grow exponentially.

### 3.1.18 Decision areas

#### 3.1.18.1 Sovereignty

The goal of digital sovereignty is autonomy, which is different from independence – it means acting with choice. It includes control over when and where data is stored and how it can be accessed. Sovereignty and autonomy are not binary concepts but ~~exist on~~move along a spectrum. The goal is to increase sovereignty and autonomy until a desired threshold is reached~~met~~. In that sense, the concept is similar to that of privacy.

#### 3.1.18.2 Resilience

Resilience in a data space is about the ability of the ecosystem and individual actors to continue functioning in the event of unforeseen problems.

#### 3.1.18.3 Scalability

Scalability of a data space is not about the volume of data but about the number of participants, the amount of the data assets shared, and the number of negotiated contracts.

#### 3.1.18.4 Control

In this context, a high level of control means that the entity operating the DSA can control access to the services as well as the content they provide. This is in direct contrast to sovereignty, where the control lies with the individual participant.

#### 3.1.18.5 Simplicity

Well-established technologies and architecture models are easier to deploy ~~a~~because~~s~~ implementing teams ~~should~~ have experience with them. The interaction model between participants as well as the business model of the data space are included in this category.

#### 3.1.18.6 Discoverability

Discoverability is the measure of how many steps are necessary to find the data offered in the data space. Since data asset information can~~ould~~ always be exchanged directly between participants, this measure only considers how complex a query would be to find all data assets currently offered in the data space ~~would be~~.

### 3.1.19 Decision support

As all decision areas are connected and partially work against each other, it is necessary to look at them holistically and not focus on one area. Make sure you weigh the importance of these decisions according to your business and technical needs. The technical maturity of the planned participants is an important factor. Many organizations are willing to compromise on their digital sovereignty in exchange for convenience and business value.

Many models exist in between the main three implementation designs. The following charts highlight some of the interdependencies between the decision areas for planning, implementing and operating a data space:

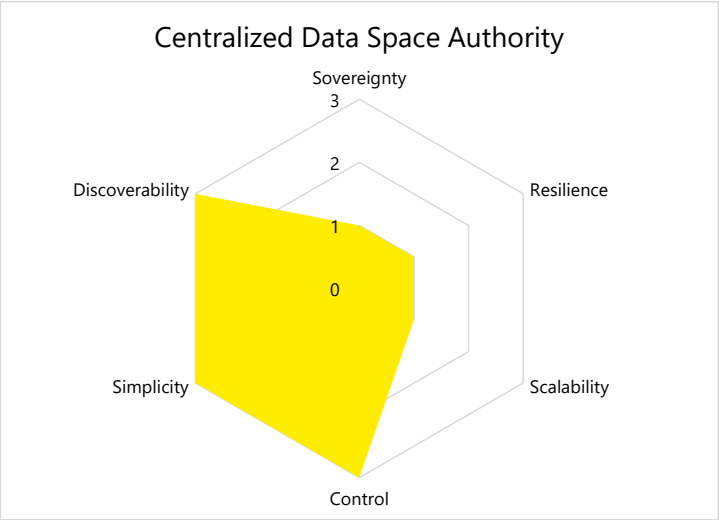



*Figure 13 Centralized data space authority*

With a centralized design the entity operating identity and catalog services has a lot of control. It is easy to setup, only one entity needs to deal with the DSA services, and participants can simply query one catalog and rely on the DSA as a trust anchor to issue a participant ID. But this design impairs participant sovereignty, is less resilient and difficult to scale as the central services will grow exponentially in their resource requirements as more participants join.

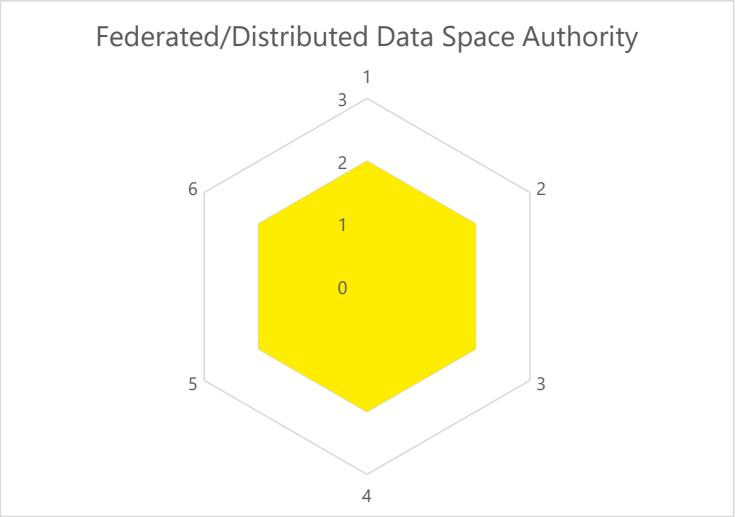



*Figure 14 Federated / Distributed data space authority*

The distributed design sits in the middle of the spectrum. Control is not exercised by a single entity but by multiple federators and thus not a single entity can make arbitrary decisions. However, participants still do not have full control over their actions, so sovereignty is still impaired. Resilience and scalability are improved by having multiple nodes of the data space services that can either be setup as partitions or as replicas. Discoverability must take into account the partitioning of the catalog and might become more complex.

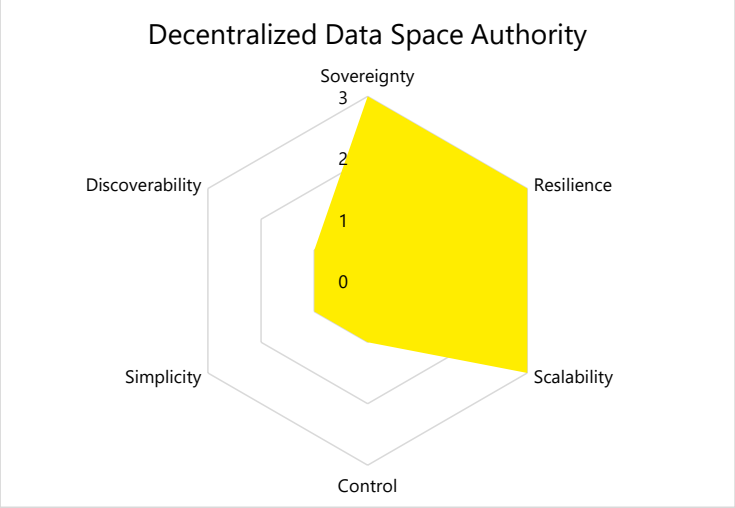



*Figure 15 Decentralized data space authority*

The aim of the decentralized design is to maximize the sovereignty of individual participants and grant them as much autonomy as possible. This reduction in dependency on central services automatically leads to higher resilience and better scalability. However, it adds complexity for the individual participant, as all participants now need to operate service nodes that participate in the discovery process of available data. Some data spaces might require additional control over participants and their actions, which is harder to achieve in a decentralized implementation.

The figure below gives a comprehensive overview of the values within the decision areas when implementing a centralized, federated/distributed, or decentralized approach.

Another way to compare the features and capabilities of the different designs is to separate the decision areas into a business and a technical perspective. Which design benefits the business value of the data space vs. which design aspects are a technical necessity? A careful compromise design-decision can be voted on by the founding parties of the data space to reach the optimal implementation.

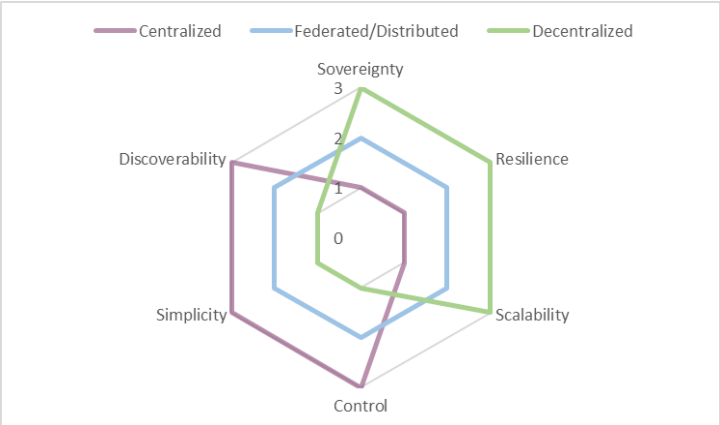
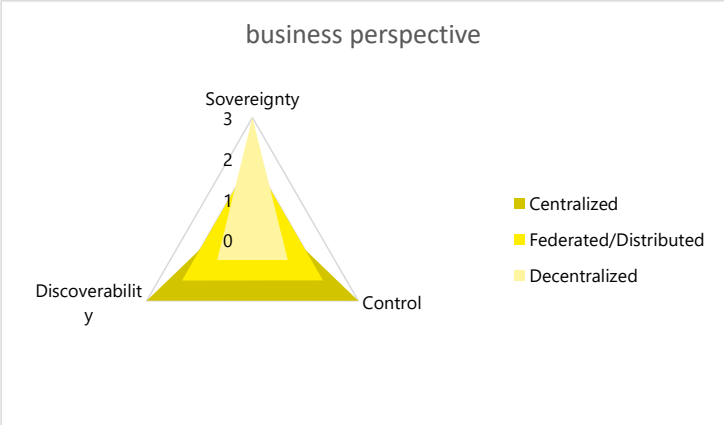



*Figure 16 Comparison of data space authority variants*




These three models are just examples of possible implementation designs. Every data space should be tailored to the needs of its participants. Any entity that wishes to participate in a data space should investigate the implementation design in detail to ensure the design grants them the aspired level of sovereignty and supports its business goals.

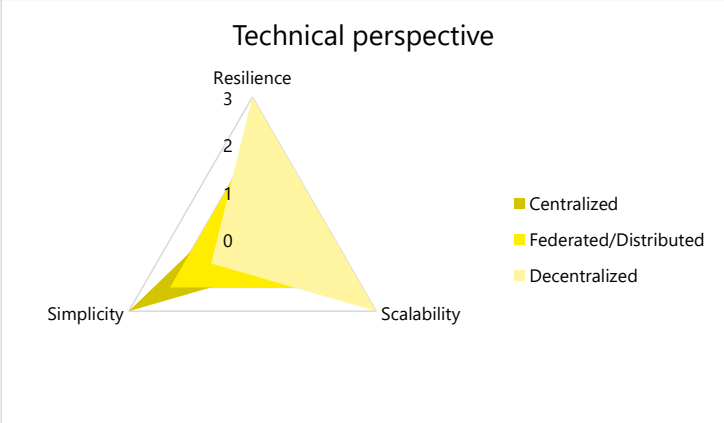



*Figure 18 Technical Perspective*

# 4 Technical agreements

Commented [SS35]: Basically reference to RAM and other tech docs

Commented [SS36]: Sebastian Steinbuss as maintainer

This section of the rulebook describes the technical arrangements required to implement an IDS-based data room. The IDS rulebook specifies what is mandatory and what is optional to implement but keep some freedom ~~on~~ how to realize these concepts (see also the section on the goals of IDS in the IDS RAM).

The technical agreements of the IDS-framework consist of the Reference Architecture Model (RAM) that provides a technology-independent perspective and the technology-specific specification on IDS-G. The two provide guidance to create the required components. The certification scheme including the certification criteria and the IDS-testbed helps validate compliance with the RAM and the specification. This is IDSA's ~~the~~ so-called magic triangle, ~~of IDSA~~ which is extended with the portfolio of open-source building blocks, such as ~~like~~ commercial solutions that are certified but not mandatorily available as FOSS. The rulebook itself provides a frame for the magic triangle by describing the overarching concept of data spaces.

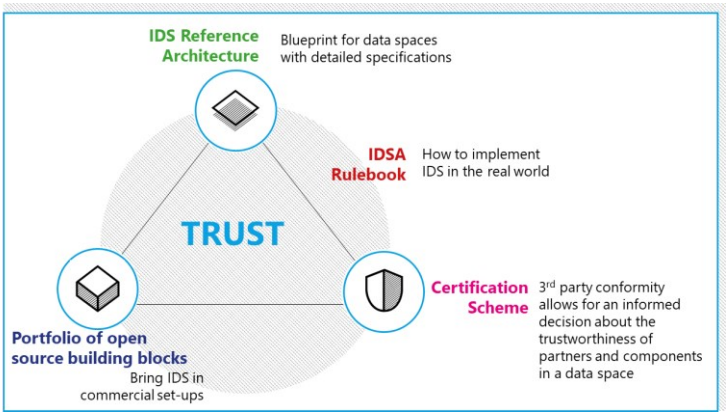



*Figure 19 IDSA Magic Triangle*

The mentioned IDSA assets have a defined release time to ensure consistency between them. In general, an IDS asset can be released after ~~the~~ approval by the IDSA working groups and ~~the~~ final approval by the technical steering committee. To achieve reliability for industrial use of the IDS assets, major releases that contain fundamental changes may ~~can~~ be conducted once per year. For m~~M~~ore details see ~~can be found in~~ the table below.

| Asset | Major releases | Approving body |
|---|---|---|
| IDS-RAM | Second quarter of a year | Working group architecture |
| IDS-G specifications | Fourth quarter of a year | Working group architecture |
| Certification scheme | Second quarter of a year | Working group certification |

| | | |
|---|---|---|
| IDS-testbed | Fourth quarter of a year | Working group certification |
| IDSA rulebook | Third quarter of a year | Working group rulebook |

## 4.1 IDS Reference Architecture Model (RAM)

Data sharing is essential for data-driven business ~~-~~ecosystems, as is the need for data sovereignty. The IDS Reference Architecture Model (IDS-RAM) defines fundamental concepts for sovereign data sharing. The IDS-RAM focuses on the general concepts, functions, and processes involved in creating a secure network of trusted data. It resides at a higher abstraction level than common architecture models of concrete software solutions. The document provides an overview supplemented by dedicated architecture specifications that define the individual components of the IDS.

The model consists of five layers: The business layer specifies the different roles that the participants can assume, and it specifies the main activities and interactions connected with each of these roles. The functional layer defines the functional requirements of the IDS, plus the concrete features to be derived from them. The process layer specifies the interactions between the different components of the IDS. It provides a dynamic view of the RAM. The information layer defines a conceptual model that ~~uses data linkage principles to~~ describes both the static and the dynamic aspects of the IDS constituents using data linkage principles. The system layer ~~is concerned~~addresses ~~with~~ the decomposition of the logical software components, considering aspects such as integration, configuration, deployment, and extensibility of these components.

Across all five layers, three perspectives need to be implemented: security, certification, and governance. The security perspective defines the common security measures for the IDS and the concepts for data usage control. The certification perspective describes the IDS Certification scheme as a foundation in the IDS. The governance perspective describes the responsibilities of ~~the~~ roles in the IDS.

The current version of the IDS-RAM that ~~is the foundation~~forms the basis for this rulebook is V4[9].

## 4.2 IDS specifications on IDS-G

Commented [SS37]: Update to current state Sebastian

IDS-G ~~is intended to~~ provides specifications and further documentation from IDSA to the public. While the RAM is technology independent, the specifications on IDS-G describe the binding of the RAM to technological concepts and focus on documentation and specifications for IDS based solutions. IDS-G's master branch is stable and therefore the reliable foundation for the development and maintenance of IDS-based solutions. It is maintained under the umbrella of the IDSA technical steering committee.

Additionally, IDS-G provides access to the IDSA open ~~-~~source projects. Currently, the following open ~~-~~source projects are available:

- IDS information model

[9] https://docs.internationaldataspaces.org/ids-ram-4/

More open -source projects will be set up by the IDSA technical steering committee in the future.

The specifications in IDS-G distinguish between four different aspects:

- **Components**: The framework for implementing IDS components as derived from the business layer in the RAM and described in the system layer, including the use of certain technologies and standards.
- **Communication**: The interaction and communication of the IDS components requires a clear specification to achieve interoperability. The communication section distinguishes between messages and message types and the interaction sequences between the components and related state machines to keep the interaction synchronized. Based on these two aspects bindings to technologies are derived.
- **Information model**: The IDS information model provides fundamental concepts to describe data products based on the IDS core concepts and fundamental standards DCAT for data assets and ODRL for contract policies.
- **Usage control**: Usage control is a fundamental mechanism in IDS. This section describes the usage contracts and how they can be realized in IDS Connectors.

The IDS-G specifications are available via GitHub[10].

## 4.3 IDS Certification

Commented [SS38]: Update

The IDS Certification is a perspective in the IDS-RAM and its approach is described in detail in the IDS Ccertification scheme (general structure, operational ~~structure~~structure, and maintenance of the certification criteria).

- The ccertification scheme[11] describes the operational model and ~~the~~ roles in the IDS Certification.
- The rules of procedure [12]include the formal outline of organizational processes
- Approval of evaluators[13]

Formatted: Bulleted + Level: 1 + Aligned at: 0,63 cm + Indent at: 1,27 cm

Formatted: Bulleted + Level: 1 + Aligned at: 0,63 cm + Indent at: 1,27 cm

[10] https://github.com/International-Data-Spaces-Association/IDS-G
[11] https://github.com/International-Data-Spaces-Association/IDS-RAM_4_0/tree/main/documentation/4_Perspectives_of_the_Reference_Architecture_Model/4_2_Certification_Perspective/CertificationScheme
[12] https://github.com/International-Data-Spaces-Association/IDS-RAM_4_0/tree/main/documentation/4_Perspectives_of_the_Reference_Architecture_Model/4_2_Certification_Perspective/RulesOfProcedure
[13] https://github.com/International-Data-Spaces-Association/IDS-RAM_4_0/tree/main/documentation/4_Perspectives_of_the_Reference_Architecture_Model/4_2_Certification_Perspective/ApprovalScheme

- ~~Conduction~~ Execution of evaluations
- The certification criteria[14] list the formal aspects of evaluations for the core components and the operational environment.

While the certification scheme and the documents listed above describe the formal aspects of IDS Certifications, the IDS testbed provides the tools and technological basis for evaluating the IDS core components.

## 4.4 IDS testbed (interoperability test)

Commented [SS39]: Update to current state (Sebastian)

Evaluation facilities for components conduct the evaluations that ensure a correct implementation of the IDS specifications and an adequate level of security in the components. Ensuring a comparable quality of all evaluations is necessary to make the certification reliable with its different security and assurance levels ~~reliable~~.

This includes:

- All evaluation facilities conduct transparent conformance tests in the „IDS reference testbed“[15] based on the regulations from the certification working group and approved by the IDSA technical steering committee.

Formatted: Bulleted + Level: 1 + Aligned at: 0,63 cm + Indent at: 1,27 cm

- All evaluation facilities assess ~~the fulfillment~~compliance with ~~of t~~he security requirements listed in the IDS criteria catalog based on test ~~specification~~s derived from the criteria. Tests that can be conducted automatically are part of the test suite[16] of the IDS-testbed.
- The evaluation facilities ~~issues~~issue a certificate when~~if~~ conformance and security tests are~~have been~~ passed.
- To ensure that the evaluation facilities conduct the evaluations according to the specifications, the certification body must assess their competence.

Ensuring interoperability between the components is one important aspect of the evaluation and covered by the ~~included~~ test suite provided.

Commented [SS40]: Update to current state Sebastian

[14] https://internationaldataspaces.org/publications/white-papers/
[15] https://github.com/International-Data-Spaces-Association/IDS-testbed
[16] https://github.com/International-Data-Spaces-Association/IDS-testbed/tree/master/Testsuite

# 5 Organizational agreements

## 5.1 Certification

Commented [SS41]: Giulia takes care

Commented [GG42R41]: Considering that this chapter is supposed to tackle all the operational challenges for data spaces, in this specific section I decided to introduce certification, in general, as an operational requirement to realize the the attribute-based trust mechanism mentioned in Chapter 3.
I then provided more details on IDS Certification as an example for such Trust Frameworks.
We can now smoothly proceed adding a mention to other types of Trust Frameworks, like the Gaia-X Labels.

@Sebastian Steinbuss, please let me know if this is in line with what you expected for IDS Certification or if more details are needed.

@PeterKoen please check my explanations of the concepts from Chapter 3.

Commented [SS43]: To be added by end of the week

Trust is *the* essential element in data spaces to overcome the reluctance to share data for fear of misuse and security concerns.

~~Chapter 3: ~~Functional requirements are an element of trust and are investigated from the functional perspective, clarifying ~~its ~~responsibilities and mechanisms in Chapter 3. This chapter discusses the operational implications~~,~~ using IDS Certification as an example.

Chapter 3 mentions two important aspects~~:~~ The first is the data space authority (DSA), which ensures trust in a data space. The second is the system enabling it, the attribute-based trust mechanism, which is based on the fundamental concepts of ~~the ~~trust anchor and trust framework. The first term refers to the entity that issues certifications about an attribute, the second to the ~~set of ~~rules imposed by the trust anchor to comply with its policies in order to be eligible for its attribute verification. Deciding which trust anchors and trust frameworks and, therefore which rules and procedures to use for issuing and validating attributes, is the task of the data space authority.

Based on the trust framework(s) selected, each data space specifies the minimum set of attributes that a participant must meet to be considered a trusted party (see also the data space self-description mentioned in Chapter 3). Based on this, each new potential member has to provide these attributes in its participant self-description ~~in order ~~to be accepted.

The DSSD must also contain clear information on which trust anchors and trust frameworks are acceptable as roots of trust within the data space, so a potential participant can decide whether to trust the data space and its members.

**The example of IDS Certification**

For the scenario described above, the IDS Certification Scheme developed by the IDSA is one available trust framework.

The trust anchor of this framework is called certification body and is a neutral party issuing certification for specific attributes. The responsibility for the certification body is taken on by a part of the IDSA head office and by additional experts hired specifically for this purpose. There are two attributes in the IDS Certification trust framework: component certification and operational environment certification.

Component certification concerns all components described in the IDS-RAM, both essential and non-essential, and ensures their required functionality and security. Operational environment certification refers to the trustworthiness of the physical environment in which the components run, as well as the processes and organizational rules there.

Both types of certifications have different options to meet the data sharing needs of companies. These options refer to the trust levels, which reflect the extent of functionalities and requirements

covered, and to the assurance levels, which refers to the method to evaluate compliance. The simplest assurance levels ares ~~is~~ based on a self-assessment mechanism, while the more advanced assurance levelsIs~~s~~ require a third-party assessment of components or operational environments. This ~~third-party~~third-party compliance check is performed by the evaluation facilities, which are specifically approved to offer this service. The approval ~~is based on a specific~~ process is defined by the IDSA certification working group.

All the details on the IDS Certification scheme, the trust ~~levels~~ and assurance levels for component certification and operational certification, the certification criteria, and the process to approve the evaluation facilities are provided in Chapter 2.

## 5.2 Running data space instances

### 5.2.1 Intra- and inter data space instance governance

It mustis ~~to~~ be recognized that the role of IDS and IDSA is to fulfill (1) within the broader landscape of existing data sharing initiatives and (2) the ambition of a federation of interoperable data spaces. This implies that IDSA ~~must~~ considers its development and deployment initiatives in the broader context of these two areas:

The striving for interoperability within a data space (intra) so that IDS can ~~be embedded and~~ provide a gradual migration path within a data space instance or data sharing initiative, and preparing for interoperability between multiple data space instances or data sharing initiatives (inter) to pave the way for the federation of interoperable data spaces - as pursued by the European data strategy.

The Data Governance Act [2] also recommends both an intra data space (domain) governance authority and an inter data space (central) governance authority. As cited from [6],

'The recently proposed Data Governance Act [2] confirms the notion of a governance structure constituted by multiple entities. For European data spaces, it is recommended to have a (domain) governance authority for each data space and a central governance authority overseeing all aspects in connection with interoperability of data spaces, i.e., the de-facto 'soft infrastructure'. This central authority will interact with all data space specific authorities.'

As noted above, IDSA and the main IDS-stakeholders playhave ~~to play~~ by working together ~~jointly providing~~to ensure ~~governance for~~ the developmenting and deployment of~~ing~~ data space instances in the broader context of introducing new or migrating/ developing existing data sharing initiatives, (intra data space interoperability), and embedding them in the European ambition of a federation of interoperable data spaces (inter data space interoperability).

**Commented [K(44]:** also link to: 4.3.8 Data Space Instances - IDS-RAM 4 (internationaldataspaces.org)

**Commented [M(45]:** Do we consider the GAIA-X Reference Model?
5. GAIA-X Operating Model Architecture document - Gaia-x - DRAFT version 1702083

**Commented [K(46R45]:** ? How does GAIA-X relate to this topc?

**Commented [P(47R45]:** No, as Gaia-X assumes that Dataspaces are a self-contained thing running on top of Gaia-X federations.

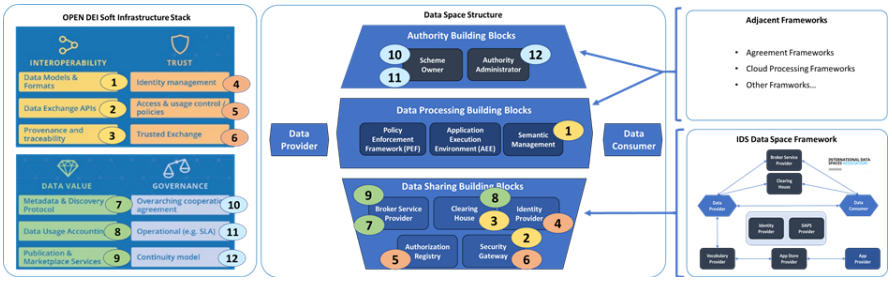



*Figure 20 Relationship of OpenDEI Building Blocks and data space instances*

This structure of a data space reflects the role of the IDS data space framework in relation to the broadly used data sharing agreement framework that is emerging as the cloud processing framework, as shown on the right ~~side~~ in Figure 1:

An approach to systematically address the interoperability challenges is provided by the new European interoperability framework ~~as~~ developed by the European Commission [8] and shown in Figure 21~~Figure 21~~.

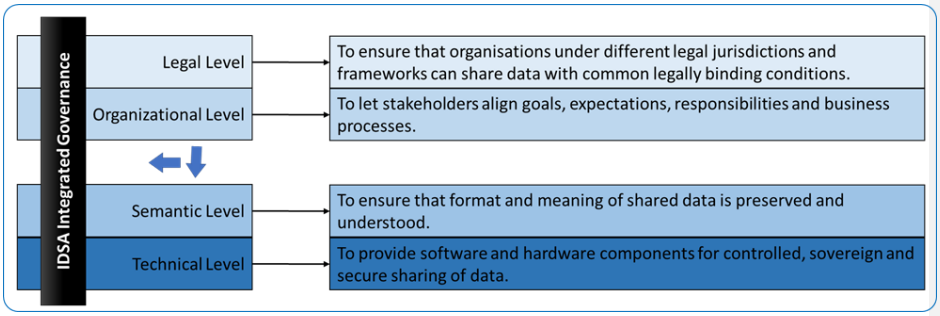



*Figure 21 Layered functional model as aligned with the New European Interoperability Framework [8].*

As Figure 21~~Figure 21~~ shows, the framework distinguishes four functional levels under an overarching integrated governance approach:

- *Technical level,* to provide software and hardware components for controlled, sovereign and secure sharing of data.
- *Semantic level,* to ensure that format and meaning of shared data is preserved and understood.

- *Organizational level,* to let stakeholders align goals, expectations, responsibilities, and business processes.
- *Legal level,* to make sure that organizations under different legal jurisdictions and frameworks can share data with common legally binding conditions.

For the governance on the identified topics, a distinction is made on their applicability to two development lines for data space instances:

- *Intra data space interoperability,* between the data space authority, processing and data sharing building blocks within a single data space instance.
- *Inter data space interoperability,* between multiple data space instances at each of the functional levels as distinguished in the framework shown in Figure 2.

To enable interoperability between data spaces, each of the functional levels shown in Figure 2 contains topics that require adequate governance. For each of the levels, these topics are identified in the following paragraphs subsequently. Their governance aspects are addressed in the next chapter.

### 5.2.2 Technical level

The technical level covers the software and hardware components for controlled, sovereign and secure sharing of data. It consists of five sub-levels with the topics that require adequate governance:

### 5.2.3 Governance instruments

The various governance instruments that may be considered by the IDSA are listed in the Table below. These governance tools each describe a different aspect of a certain activity and are used in the next section to provide a multidimensional overview of the intra- and inter-data spaces governance.

**Table 2: IDSA governance instruments**

| Governance Instrument | Governance instrument description |
|---|---|
| *Standardization* | Ensuring that the tasks, processes, and guidelines around this activity are formalized, documented, and aligned between data spaces instances and the IDSA. The standardization efforts can generally be used as a blueprint or starting point for the stakeholders. |
| *Certification* | Validating that stakeholders act according to the standardized way of working. Certification is divided into component certification (certification of |

| | |
|---|---|
| | technical/software components) and organizational certification (organizational and legal processes). |
| *Development* | The activity may require (software) development tasks as part of the realization, which should be compliant with the standardization activities and best followed by a certification. |
| *Operations* | The operations are about the exploitation and usage of the developed components. The operations can be certified as part of the organizational certification. |
| *Communication* | The dissemination of the activity is an important aspect which might include the awareness of the standardization and certification, but also contains marketing aspects. |
| *Support* | The support activities include the structured assistance from stakeholders involved in the operation, development, certification, and communication activities. |

### 5.2.4 Governance for inter data space interoperability

There will be no single data space. Individual sectors or communities are expected to develop their own data space instances. Being able to seamlessly share data over these data space instances brings clear advantages. It extends the reach and scope of accessible data and allows the development of new business models and services across sectors and regions. Therefore, interoperability between data space instances adds major value, resulting in a federation of interoperable data spaces as shown on the right of Figure 3.

#### 5.2.4.1 Interoperability architecture considerations

##### 5.2.4.1.1 Harmonization

The Data Sharing Coalition is an open and growing international initiative in which, a large variety of organizations are collaborating to drive cross-domain data sharing at scale. Its results on cross-domain data sharing were recently published in the "Data Sharing Canvas" [7].

The Data Sharing Canvas compared various harmonization options. Full harmonization of data spaces, in which existing data spaces adjust their implementations to follow a common cross data space design, is the ideal solution to achieve multilateral interoperability. However, it impacts all data space participants, requiring significant investments and will therefore not be adopted. Bilateral harmonization of data spaces, in which individual data spaces organize custom interoperability bilaterally depend on individual participants implementing specific harmonized solutions and will therefore limits large scale data sharing cross data space. As an alternative partial harmonization,

which introduces the new role "data space proxy", overcomes these limitations of full and bilateral harmonization. The proxy absorbs the complexity of harmonization for data spaces and its participants as much as possible by implementing all harmonization requirements. This enables a data provider in one data space to share data with a data consumer in another data space, while limiting the impact on existing data providers and data consumers.

The main function of the proxies is to translate data space specific transactions to their harmonized equivalents:

- Proxies translate data space specific language into a harmonized language in the harmonization domain to enable multilateral end-to-end interoperability,
- Proxies facilitate trust across data spaces by conforming to the rules and agreements of the trust framework,
- Proxies enable the discovery of data providers across data spaces.

The proxies implemented by all data spaces form a network, the harmonization domain, which enables each data space to share data effortlessly with other data spaces.

#### 5.2.4.1.2 Interaction topologies

To ensure interoperability between data space instances, the intermediary roles of these will have to interact and to exchange (meta)data. These interactions may be designed through various metadata role interaction topologies (MRIT). The decision to implement a specific type of MRIT results in a governance role for the IDSA and other IDS stakeholders in their development or deployment. They are shown in Figure 6.

The two basic types of MRITs applicable for exchanging (meta)data between intermediary roles from different data space instances, as shown in the figure:

- *Intermediary-to-Intermediary MRIT (I2I-MRIT),* where the exchange of (meta)data between the intermediary roles is done on a bilateral, peer-to-peer basis.

An example is the broker service provider that bilaterally exchange (meta)data on available data sources.

- *Bridged MRIT (B-MRIT),* where the exchange of (meta)data between the intermediary roles uses an overarching bridging function.

An example of this is the legal framework with associated participant registry for managing that data spaces instances and its participants, which legally adhere to an overarching legal framework.

**Formatted:** Bulleted + Level: 1 + Aligned at: 0,63 cm + Indent at: 1,27 cm

**Commented [P(48]:** This is an approach that conflicts with the IDS Protocol and current thinking on Dataspace Interoperability. Proxies infringe on the autonomy and sovereignty of individual participants and thus are NOT a desired implementation pattern.
I would recommend to remove this section from the current RuleBook and provide an update on Dataspace Interoperability & Interconnectivity in the next version.

**Commented [TM49R48]:** To be removed

**Commented [TM50R48]:** … and sync with upcoming publication (Sebastian, Pieter)

**Formatted:** Bulleted + Level: 1 + Aligned at: 0,63 cm + Indent at: 1,27 cm

**Formatted:** Bulleted + Level: 1 + Aligned at: 0,63 cm + Indent at: 1,27 cm

For interoperability of the various data space capabilities, different interaction topologies may be most suitable, requiring a different governance role for IDSA in developing or deploying these topologies.

#### 5.2.4.2 Concluding on the intra data space development

The exact relationship between data spaces and the IDSA may vary from data space to data space, but generally the IDSA can provide guidelines, frameworks, or policies for how processes are structured within a data space instance. The data space itself is responsible for executing this process.

A complete set of APIs for interacting with the various building blocks and capabilities of the data space structure (as shown in Figure 1) should be identified and defined. This will be done under the responsibility of the IDSA technical steering committee (IDSA-TSC) and reported on within the future releases of the IDSA Rule Book [11]. This may go beyond the current scope of APIs defined by the IDSA. Specific attention should be given to:

- An API for the policy execution framework (PEF) to be used by data apps for using the PEF's data control and sovereignty features, as identified for the application execution environment (AEE) in Table 4.
- APIs for managing and accessing data space membership identities (as provided by the DAPS and ParIS functions) to be used in the authorization flows for individual data transactions, as identified for the data space membership (DAPS, ParIS) in Table 4.
- An API for accessing (cloud) processing capabilities for locally executing data apps, as identified for cloud integration (GAIA-X) in Table 4.

**Formatted:** Bulleted + Level: 1 + Aligned at: 0,63 cm + Indent at: 1,27 cm

An adequate and future-proof governance process for managing the IDSA standards is needed. The IDSA standards need management and maintenance, including further development with backward compatibility of versions. This process will be further explained in the future versions of the IDSA rulebook [11].

**Commented [P(51]:** This information is outdated as it's superseded by the current IDS protocol development.
Also it is at a level that should rather be described in RAM than in the RuleBook. I'd recommend removing this section.

#### 5.2.4.3 Concluding on the inter data space development

The Data Sharing Coalition [7] suggests the proxy model in combination with a harmonization domain and protocols as the architecture to enable interoperability between data space instances for inter data space interoperability as shown in Figure 5. As the need for inter data space interoperability is rapidly growing, the IDSA must assess how the development of (IDS-based) data spaces and the role of IDSA in co-developing the data space proxy interfaces and harmonization protocols are impacted. The IDSA governance model of the inter data space architectures and standards is to be defined.

An overarching data space scheme defining and implementing joint legal and operational agreements between adhering data spaces instances will be central to achieve this interoperability. It provides a legal framework to which individual data space instances (and their subscribers) agree to adhere.

The role of IDSA as the overarching data space scheme owner should be assessed and adequate governance be provided. The IDSA may (have to) perform an ongoing operational task in fulfilling this role.

The need for inter data space interoperability has been identified as relevant for various European research and development initiatives. For instance, the Data Sharing Coalition [7] and the OPEN DEI initiative (in the future releases of their 'Design Principles for Data Spaces' [6]) are expected to explicitly address this topic. An active role of IDSA and its stakeholders should be pursued in these European initiatives.

Commented [AS(52]: A few things in these paragraphs are said twice and redundant. If you want me to rewrite I am happy to do so! Some paragraphs I just deleted because they where identical to others

### 5.2.5 General approach

Interoperability within and across AI data spaces: Development lines of AI data spaces provide the building blocks for managing trust, data sovereignty and agreements to share data and algorithms - to execute AI algorithms and data apps. Given the European ambition to federate European data spaces, both single and multiple data spaces require adequate governance to realize interoperability within and across data spaces. Therefore, a distinction is made between two development lines for data spaces:

• Intra data space interoperability, between the various building blocks within an individual AI data space instance. The definition of federation ("A change from one central data powerhouse to democratization of data" NLAIC) indicates that individual AI data space instances have a high degree of autonomy for their own internal agreements and ICT landscape. Intra data space interoperability provides a reference architecture based on common building blocks and path for developing AI data space instances in an efficient and aligned manner, providing a rich set of features to support AI challenges and requirements. It leaves individual data spaces the option to deviate internally from the reference architecture.

• Inter data space interoperability, between multiple data space instances. Interoperability between AI data space instances is key for the federation of AI data spaces to seamlessly interconnect. This is the goal of the NLAIC data sharing working group in realizing a cross-sectoral data sharing infrastructure for AI and aligns with the EU data strategy. Inter data space interoperability requires prescriptive guidelines for individual data space instances to ensure interoperability between them.

### 5.2.6 Conclusions

The European OPEN DEI endeavor (let by the European Commission) aimed at defining the building blocks and standards for data spaces, to realize interoperability between the building blocks within specific data space instances (intra data space interoperability) and between various data space instances (inter data space interoperability).

The important role of IDSA and the IDS-stakeholders are to provide the governance for data space instances in the broader context of:

-the introduction of new or migration/ evolution of existing data sharing initiatives, intra data space interoperability, and

-the embedding thereof within the European ambition of a federation of interoperable data spaces inter data space interoperability.

# 6 Legal dimension

This section of the Rulebook gives an overview about the regulatory framework and describes IDSA's approach regarding compliance with regulatory requirements and contractual agreements.

## 7 6.1 Regulatory Framework

The lack of a general legal status (access regime) for data, partial application of IP rights and trade secret protection and the restrictions of personal data protection regime overall give rise to a fragmented and incomplete regulatory framework. With a view to address these current shortcomings regarding the sharing and reuse of data, in February 2020 the EU Commission communicated the "European strategy for data" describing the vision of a common European data space. As part of EU's digital strategy the Commission has proposed different regulations on harmonised rules for data governance, data access and use:

Beside other regulations on digital topics[1], the Data Governance Act (DGA)[2] entered into force on 23 June 2022 and, following a 15-month grace period, will be applicable from September 2023. On 23 February 2022, the Commission proposed a Regulation on harmonised rules on fair access to and use of data, the Data Act Proposal (DA-E)[3].

Field Code Changed

Field Code Changed

Field Code Changed

With both Acts the Commission aims to make more data available for use, and set up rules on who can use and access what data for which purposes across all economic sectors in the EU:

The DGA aims to make more data available by regulating the re-use of publicly/held, protected data, by boosting data sharing through the regulation of novel data intermediaries and by encouraging the sharing of data for altruistic purposes.[4] It aims to make public sector data available for local businesses, researchers and communities for the development of innovative data-driven services at a larger scale. It has a specific focus on the public sector data which is subject to legal restrictions and thus left out of the scope of the Open Data Directive. Therefore, the proposal covers public sector data which is legally protected on the grounds of: (a) commercial confidentiality including the trade secrets; (b) statistical confidentiality; (c) intellectual property rights of third parties; (d) protection of personal data. This objective of providing access to data that is not accessible as "open data" may be seen as indicative of the emergence of a distinct regime for the data held by public bodies. The public sector bodies enabling the use of such protected data are required to be technically equipped to ensure that data privacy and confidentiality are fully preserved. The proposal does not interfere with the substantive rights on data as it refrains from prescribing a right of access or reuse but lays out certain harmonized rules and conditions guiding Member States for establishing mechanisms for the reuse of publicly held data.

Field Code Changed

The DA-E aims to ensure fairness in the digital environment, stimulate a competitive data market, open opportunities for data-driven innovation and make data more accessible for all

by providing consumers and businesses access to the data of their devices.[5] The DA-E is regarded to be an essential building block of the European data spaces. It is guided by the understanding that B2B contractual agreements do not fully guarantee adequate access to data for SMEs or start-ups—entailing a contractual framework providing clarity as to the rights and remedies regarding accessing, processing, sharing and storing of data in order to limit the misuse of such data. The proposal acknowledges the importance of a harmonised data governance regime in achieving competitiveness, innovation and sustainable growth in all sectors and making the Union's transition to a green digital economy a success. The proposal introduces interventions to the current legal landscape of B2B data sharing and access in two dimensions: first, contracts as voluntary agreements and second, statutory access rights or obligations to make data available together with the general rules to be complied while performing these obligations or exercising the rights.

[1]https://digital-strategy.ec.europa.eu/en/policies/strategy-data[1] Digital Markets Act (DMA) https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM%3A2020%3A842%3AFIN; Digital Services Act (DSA) https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020PC0825; AI Act https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A52021PC0206.

[2] REGULATION (EU) 2022/868 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act); https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R0868.

[3] Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on harmonised rules on fair access to and use of data (Data Act); https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2022%3A68%3AFIN.

[4] https://digital-strategy.ec.europa.eu/en/policies/data-governance-act-explained.

[5] https://ec.europa.eu/commission/presscorner/detail/en/ip_22_1113.

Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed

Beside these specific Acts, further legal topics needs to be considered when sharing data, including antitrust/competition, Data Protection and Security, Copyright, patents/Intellectual property. The regulatory development might have more impact on the concept and operationalization of data spaces in the future and needs to be monitored to ensure compliance.

The operationalisation of data governance and the establishment of data spaces require a robust methodology both to navigate through the existing regulatory patchwork (scattered in various legal instruments) and also to implement the upcoming legislative agenda of the EU. Providing guidance to address concrete problems in a future-proof manner entails an assessment and combination of various regulatory tools, contractual models, design principles, and organisational structures. To this end, the below ***four-pillar data governance framework*** , as a theoretical and conceptual scaffold, draws up a "legal anatomy" of data governance consisting of:

i. The substantive rights and obligations pertinent to data transactions (rights on data);
ii. the contractual dimension;
iii. the organizational aspects; and
iv. the technical implementation.

Beside the own responsibility of participants in a decentralized organization and even though IDSA doesn't focus on the legal dimension, IDSA discusses and aligns on legal matters with other initiatives. The alignment with other initiatives regarding the legal dimension is even more important as often (and by nature) most legislations need to be translated into practical approaches and solutions and a common understanding of the legal terms is necessary to create a trustworthy and reliable EU data sharing landscape.

Therefore, IDSA has a Legal Framework Task Force to discuss regulatory developments and specific legal topics and to organize the collaboration and contribution of IDSA members regarding the legal dimension.

## 6.2 Legal Agreements & SITRA Rulebook

The analysis of the relevant legal frameworks pertaining to data transactions reveals that there exist many gaps and overlaps in the current legal landscape mostly because, i) significant parts of the data do not have a default legal status as intangible assets, and ii) these legal regimes are not designed with the needs of the data economy or the specificities of data transactions in mind.

As legislation only provides the general framework for data sharing, the legal dimension of a data space includes a contractual framework so that the different participants can agree on specified rules that fit into their data sharing context. In a decentralised organisation where participants are free to choose their contract partner and freely agree on contract terms, the contractual framework means a suggested model of terms that can be amended according to the specific needs (template approach).

Considering IDSA's focus on other dimensions as well as considering the importance of alliance with other initiatives, IDSA follows an "adopting & consolidation approach" regarding contractual framework. Therefore, IDSA did not invent an own "new" set of legal agreements for IDS participants to be used but decided on suggesting an already established contractual framework modified regarding IDS specifics.

**Commented [Sö53]:** Is this a proper description ? Please feel free to modify.

**Commented [SS54R53]:** Sebastian will add some sentences on this, What is the IDS Focus and what i our context how we relate to other initaitves.

SITRA's Rulebook for a Fair Data Economy provides tools for networks where companies and other organizations can share data and create new services. The Rulebook model includes contractual templates and tools for building a data sharing network. It sets out legal, business, technical, security, and administrative rules as well as ethical guidelines to be observed by organizations in data sharing networks. The Rulebook model consists, among other things, of contractual templates, a set of control questions and a draft code of conduct that can be used to create a customized rulebook for a given data network. Sitra published the first version of the Rulebook for a Fair Data Economy in 2020 and it has since been updated several times. The Rulebook model is backed by Sitra's long-term work on the fair data economy and a large group of experts from companies and other organizations who have made valuable contributions to the Rulebook model.[17]

The basic principles embedded in the Sitra Rulebook fit well to the goals of this IDSA Rulebook, as well:

**Sovereignty**

IDS has set sovereignty for data owners as its key design principle. In the Rulebook Data Provider has sovereignty over its data. The instrument for the Data Provider to exercise its sovereignty is through the Dataset Terms of Use, in which the Data Provider can decide to whom grant access and to set the terms under which it releases the Dataset for use by others in the Data Network.

**Trust**

Enabling trust is the first of the IDS foundational concepts. Trust is encompassed in the Sitra Rulebook especially through the balance it builds between the sovereign Data Providers and Data Users building new business from the data. For instance,

[17] https://www.sitra.fi/en/publications/rulebook-for-a-fair-data-economy/

seizing the provision of data is allowed, the termination period can be set to fit the needs of the business or even initial fixed terms can be agreed. As the default, the Data already distributed may still be used by the Data User even after termination, but the provision of new Data will be seized. The balance is also found in the clauses defining the borderline between Data and Derived Material in a way that it fits the context and the needs of the Data Provider and the Data User. The Sitra Rulebook has also clauses on auditing and extensive tools for ensuring that data security and ethical principles are taken into account in the design of data networks.

Considering the IDS dimensions, the SITRA templates are a valuable basis to create the contract framework for data sharing based on IDS principles and specifications. IDSA follows a "narrow" approach regarding the suggested contract templates as the idea is to provide a general framework that should be amended according to the specific needs.

**Commented [SJ55]:** The meaning of this should still be opened more.

## 6.3 Contract templates for IDS

Based on the SITRA templates IDSA will start drafting additional components for contract templates for IDS (that will be published after this Rulebook). Such templates will be attached to this Rulebook and will be updated from time to time considering new developments.

The IDS contract framework will not duplicate all components of the SITRA Rulebook. The full set of SITRA Rulebook templates is intended to be used in creation and set-up of a data space, including governance models of the data ecosystem. These may not be necessary for the purposes of this IDSA Rulebook. Therefore, the IDS contract framework will focus on additional components and guidance highlighted in different use cases of data sharing implemented under the IDS specifications. These may include for instance domain specific Dataset Terms of Use –templates or more detailed components for cross-continent data sharing or privacy. In case IDS contract framework requires modifications to the SITRA Rulebook's General Terms and Conditions, they will be proposed also to Sitra's workgroup in order to maintain compatibility and to avoid different parallel versions of General Terms and Conditions.

# 8 Summary and outlook

**Commented [SS56]:** Draft will be created by Marko, Olaf, Peter, Sebastian

The IDSA Rulebook Version 2 empathizes on the growing need for structural approaches of the need for sharing and exchanging data while maintaining data sovereignty. While the topic is still under development the use of guiding principles helps to find solutions that fit to the need of this growing market. This includes the understanding of the current European regulation and legislation, but not limiting it to this, as data spaces are meant to be international and not bound to local law and regulations.

The Functional analysis of both parts, the creation of a data space and the data space authority, as well as the requirements and obligations of a participant in a data space is therefore a central piece of this document. The comprehensive outline provides guidance to the creation of data spaces. This is complemented by the governance view on the rights and obligations for data spaces in a whole and the data space instances.

The analysis of the legal framework for data spaces is still ongoing and will be subject to continuous debate and will be published along with this rulebook in the future.

The endeavor of IDSA and its partners in data spaces is supported by the Data Spaces Support Centre. This project will provide additional guidance.

Based on the recent IDSA work, additional publications will provide more insights into the field of data spaces. The Data Space Landscape document and reports on technical and semantic interoperability will be part of the future work for this Rulebook. The relationship of the different stakeholders in the data space landscape and how they form a comprehensive framework will be further investigated.