Knowledge Base

Export

International Data Spaces Association

2026-04-15

Contents

1 Knowledge Base

Version: 20260415-74-12a8bf9

Generated: $(date -u +‘%Y-%m-%d’)

2 Knowledge Base

Version: 20260415-74-12a8bf9
Generated: 2026-04-15

3 Home

4 Welcome to the Knowledge Base

The IDSA Knowledge Base serves as a comprehensive repository of information and resources developed by the IDSA Working Groups. This platform brings together a collection of approved and published deliverables, offering valuable insights and guidance for the general public. While the Working Groups are continually refining and updating their materials, the Knowledge Base features only the finalized versions that have undergone review and approval. As a result, the documents included may not always reflect the most current status of ongoing drafts, but they represent the authoritative and officially released content.

4.1 What you will find in the IDSA Knowledge Base

- **Manifesto of International Data Spaces:** *(tbd)* Discover the foundational vision and guiding idea behind international data spaces, outlining their purpose, core values, and the future they aim to enable.
- **Principles of Dataspaces from the Rulebook:** Access the key principles and rules for trustworthy, secure, and interoperable dataspaces, as agreed and published in the official Rulebook.
- **Practical Guidance from the Reference Architecture:** Find actionable recommendations and step-by-step instructions on how to design, build, and operate data spaces, based on the approved reference architecture deliverables.
- **Glossary of Commonly Agreed Terms:** Explore a glossary featuring standardized definitions and explanations of essential terms used across international data spaces, ensuring clarity and shared understanding for all stakeholders.

  Tabs: Use the header tabs to switch between **Home / Knowledge / About**. The main content is in the Knowledge Tab

5 What is a data space?

6 What is a data space?

It is really cool, let me explain briefly and then go and read the knowledge base

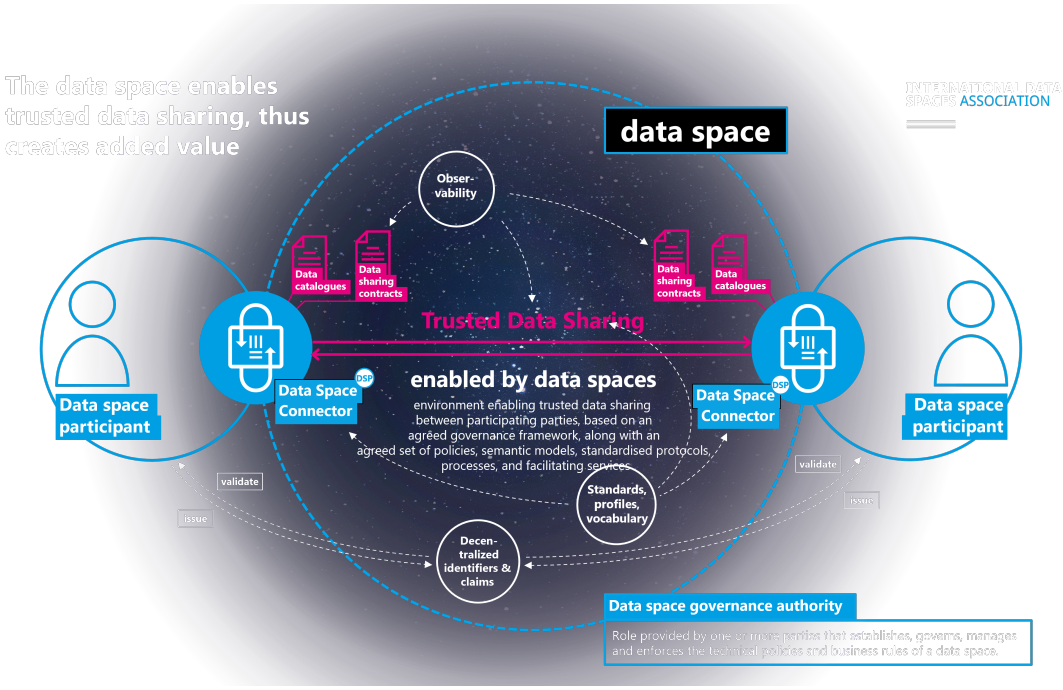

Figure 1: What is a data space

7 Cover

8 README

8.1 IDSA Rulebook

8.2 Table of contents

8.3 Purpose & Motivation

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 on how to implement or use a technical artifact 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.

8.4 How to Contribute to IDS Rulebook?

We value your feedback and encourage all interested parties to review, comment, and contribute to the ongoing development of this Rulebook. Collaboration and open communication are essential to our success, and your input is vital to refining and enhancing the Rulebook, making it a more comprehensive and valuable resource for all. You can provide feedback on IDS Rulebook by creating an issue in this repository.

One can become a contributor by joining the Rulebook Working Group as a member of International Data Spaces Association. This is possible via contacting the co-chairs listed on IDSA Website.

Please also visit our Code of Conduct and the Contibuting information.

8.5 Changelog

We maintain a Changelog to keep track of updates to the document.

8.6 License

The IDSA Rulebook is published under the Creative Commons License 4 CC-BY. ## Sources ##

Please find the resources for the IDSA Rulebook on GitHub.

9 Front Matter

10 IDSA Rulebook V2

10.1 Publisher

International Data Spaces Association

Anna-Louisa-Karsch-Str. 2

10178 Berlin

Germany

10.1.1 Editor

Sebastian Steinbuss,

International Data Spaces Association

10.2 Copyright

International Data Spaces Association,

Dortmund, Germany, 2023

Figure 2: Creative Commons License

This work is licensed under a Creative Commons Attribution 4.0 International License.

10.3 Authors and Contributors

- tbd

11 Contributing Projects

12 Contributing Projects

These projects contributed to the IDSA Rulebook.

Thank you!

12.1 Industrial Data Space

Figure 3: Industrial Data Space

Industrial Data Space

12.2 Industrial Data Space +

Figure 4: Industrial Data Space

Industrial Data Space

12.3 Industrial Data Space 3

Figure 5: Industrial Data Space

Industrial Data Space

12.4 BOOST 4.0

BOOST 4.0

12.5 AMable

AMable

12.6 MIDIH

Manufacturing Industry Digital Innovation Hubs

Figure 6: Boost 4.0

Figure 7: AMable

Figure 8: MIDIH

Figure 9: Research Center Data Spaces

12.7 Research Center Data Spaces

Research Center Data Spaces

12.8 DEMAND

Figure 10: DEMAND

DEMAND

12.9 MUSKETEER

Figure 11: MUSKETEER

MUSKETEER

12.10 AI4EU

AI4EU

Figure 12: AI4EU

AI4EU

12.11 Dataspace Mobility

Dataspace Mobility

12.12 Qua4lity

Qua4lity

Figure 13: Dataspace Mobility

Qua4liy

Figure 14: Qua4liy

12.13 MARKET 4.0

Figure 15: MARKET 4.0

MARKET 4.0

12.14 OpenDEI

OpenDEI

12.15 EUR3KA

EUR3KA

12.16 DIH^2

DIH^2

12.17 AI Marketplace

AI Marketplace

12.18 SPEAKER

SPEAKER

12.19 EuHubs4Data

EuHubs4Data

Figure 16: OpenDEI

Figure 17: EUR3KA

DIH2

Figure 18: DIH2

Figure 19: AI Marketplace

Figure 20: SPEAKER

Figure 21: EuHubs4Data

Figure 22: Level-Up

12.20 Level-Up

Level-Up

12.21 TRUSTS

Figure 23: TRUSTS

TRUSTS

12.22 FlexiGoBots

Figure 24: FlexiGroBots

FlexiGroBots

12.23 BD4NRG

BD4NRG

12.24 Dat4Zero

Dat4Zero

Figure 25: BD4NG

Figure 26: Dat4Zero

13 Introduction

14 Introduction

14.1 Who should read this Rulebook?

This Rulebook is addressed to the broad community of actors who design, build, operate, regulate, or participate in data spaces. That includes private enterprises, public sector organizations, research institutions, standards bodies, and individuals who are responsible for data governance, stewardship, compliance, or innovation. As data sharing assumes an ever more fundamental role in economic activity and public policy, a clear understanding of the principles, requirements and governance models set out here is essential.

The Rulebook offers practical guidance for those working with diverse forms of data sharing — from peer-to-peer sharing and federated ecosystems to data marketplaces and platform-based services. It is especially useful for readers who seek to promote trustworthy, sovereign, and legally compliant data sharing; to manage business risk and contractual governance; and to implement technical architectures that preserve participant autonomy and agency.

14.2 Goals and scope of the IDSA Rulebook

14.2.1 The purpose and scope of the IDSA Rulebook

The IDSA Rulebook supports the creation, operation, and growth of data spaces by distinguishing mandatory requirements from optional, value-adding practices. Its scope spans technical, commercial, and legal dimensions:

- Common technical guidance, including functional requirements and specifications.
- Recommendations for applying IDSA technical artefacts and for alignment with partner frameworks.
- Operational guidance for collaboration, roles, and processes that enable data space ecosys-

tems.
- Perspectives on implementing and complying with international legal and regulatory obligations to facilitate trusted, cross-border data sharing.

The Rulebook describes how technical roles (for example, Participant and Data Space Governance Authority — DSGA) relate to economic and legal responsibilities, and how these roles may map to obligations under instruments such as the EU Data Governance Act, the EU Data Act, and international programms like the Data Free Flow with Trust (DTFF)

14.2.2 Goals of the IDSA

The International Data Spaces Association (IDSA) aims to cultivate a vibrant practitioner community and to provide concrete guidance that enables the realization of data spaces across a range of capabilities and organisational models.

To that end, IDSA develops the Data Space Requirements (the IDSA Rulebook), the Reference Architecture Models (RAMs), complementary implementation and operations guidance. IDSA also engages with international standardization bodies and open-source initiatives to harmonize and share the knowledge contributed by its members, thereby supporting the global adoption and interoperability of data space technologies and business models.

The central objectives for data spaces is the establishment of trustworthiness in data sharing. The Manifesto of international data spaces articulates the fundamental principles that underpin these objectives:

- Data Spaces are a mechanism of Trust (Data Spaces enable Trusted Data Sharing)
- Your Data, Your Choice (Actors shall have full autonomy in deciding with whom they share data with and under what conditions)
- With great responsibility comes great power (Actors shall be responsible for ensuring their freedom to act autonomously)
- Data Spaces are Decentralized & Neutral (All actors shall be treated equitable in their rights and obligations)
- Data does not flow through the Dataspace (Sharing of data is executed on private channels)
- Unity in Standards - Freedom in Implementation (Data Spaces shall be based on international standards)
- There is no single platform to rule them all (Data Spaces shall be infrastructure agnostic)
- Data Spaces are not Data Ecosystems (Data Spaces are building blocks for data ecosystems)
- The opportunity is boundless (Data Spaces shall be business model agnostic)
- Act in good faith, but verify (Actors shall honor all data contracts and its associated policies and verify adherence by others)

These principles provide the foundation for trusted data sharing and for the consequent development of data-driven services and business models.

IDSA specifies foundational requirements and implementable reference architectures that enable organizations of all sizes and sectors to offer, discover, negotiate, and consume data-sharing arrangements for their digital assets.

You can find additional information about data space elements from IDSA in the following sources:

- The IDSA website (https://www.internationaldataspaces.org) provides information about our work, use cases, publications and events.
- The IDSA GitHub repositories (https://github.com/International-Data-Spaces-Association) host specifications, reference implementation guidelines and an open forum for

member collaboration via issues, discussions and pull requests.

14.3 Relationship with other organizations, projects & initiatives

14.3.1 How Do Initiatives Relate in the field of Data Spaces?

The field of Data Spaces and trusted data sharing is rapidly evolving. As industries, research institutions, and governments seek to collaborate across organizational and national boundaries, the need for interoperable approaches to data exchange has become critical. Multiple initiatives and groups are contributing to this ecosystem, each playing a distinct but complementary role. Together, they create the foundation for standardized, reliable, and scalable Data Space solutions. This section explains how these initiatives relate to one another, highlighting their contributions to specifications, standards, open-source implementations, and testing frameworks. It includes organizations, projects or standards that have made a significant contribution to or own a dependency for the IDSA Rulebook. This section will be regularly updated as the ecosystem grows and further contributions are made or when new dependencies arise.

14.3.2 International Data Spaces Association (IDSA)

The International Data Spaces Association (IDSA) brings together global members from both industry and research. Its mission is to develop and promote the concept of data spaces, covering the full spectrum from legal frameworks and business models to technology foundations.

IDSA provides a unique forum for aligning perspectives across its community. By collecting and structuring requirements, the association ensures that the needs of diverse stakeholders are represented in discussions about data space architecture. The value of this end-to-end perspective lies in its ability to integrate legal, organizational, and technical considerations into a coherent vision. At the technical level, IDSA emphasizes the importance of a common core for specification and standardization. This core is designed to foster interoperability between data space solutions at the protocol level. To achieve this, IDSA aggregates member requirements and channels them into international specification projects (e.g., within the Eclipse Foundation) and formal standardization activities (such as ISO/IEC JTC 1/SC 38 or CEN/CENELEC Joint Technical Committee (JTC) 25).

In short, IDSA serves as the bridge between conceptual discussions, community requirements, and downstream technical specifications.

14.4 ISO/IEC 20151: A Common Foundation

One of the major cornerstones in formal standardization of data spaces is **ISO/IEC 20151, Information Technology – Cloud Computing and Distributed Platforms – Dataspace Concepts and Characteristics**. This standard provides a clear and authoritative definition of Data Space concepts, distinguishing them from related ideas such as data warehouses, data lakes, data fabrics, or data meshes. By describing the essential characteristics and requirements of a data space, ISO/IEC 20151 reduces ambiguity and helps ensure consistency in design and implementation. The standard is not only conceptual; it also provides the baseline for interoperability. By establishing common ground, it enables both intra-data space (within a single ecosystem) and inter-data space (across ecosystems) technical compatibility. In doing so, it creates the foundation on which further specifications and open-source implementations can build.

14.5 Dataspace Protocol (DSP) and Decentralized Claims Protocol (DCP)

The Dataspace Protocol (DSP) and Decentralized Claims Protocol (DCP) are two key specification projects that operationalize the concepts defined by IDSA and ISO/IEC 20151.

- DSP focuses on the communication mechanisms required for trusted data sharing between participants in a data space.
- DCP addresses decentralized identity and claims management, which are central to ensuring trustworthiness and accountability.

Both protocols are developed under the governance of the Eclipse Foundation, ensuring transparent processes and adherence to rigorous intellectual property rules. While they are rooted in the requirements articulated by IDSA, their development is open to a broad community beyond the association. This open governance model fosters collaboration and ensures that the specifications can evolve in line with real-world needs.

14.6 Eclipse Dataspace Working Group (EDWG)

To coordinate and endorse data space-related efforts, the Eclipse Foundation has established the Eclipse Dataspace Working Group (EDWG). The EDWG serves several purposes:

- It associates and endorses specification projects like DSP and DCP.
- It provides a governance structure through its committees, which decide on project alignment and associations.
- It supports submissions towards ISO Publicly Available Specifications (PAS), ensuring that community-driven specifications can support international standards, speeding up standardization in response to urgent market needs.

By bringing specifications and open source implementations under one umbrella, the EDWG provides coherence and continuity. It is a key mechanism that, through the participation of its members, it ensures data space technologies remain consistent, interoperable, and aligned with global standards.

14.7 Eclipse Dataspace Components (EDC)

Specifications alone are not enough; they must be validated through implementation. This is where the Eclipse Dataspace Components (EDC) project plays a vital role. EDC is a reference implementation of both DSP and DCP. It provides a framework for developers to build data space components with a common core and extensibility mechanism. This design allows rapid integration with existing technologies, such as storage systems, vault services, event processing platforms, or policy engines. Compliance is a central focus of EDC. Each release version of the framework, together with a defined set of core extensions, is tested against a Technical Compatibility Kit (TCK). Successful test results are published openly, ensuring transparency and building trust in the framework's conformity to the specifications.

For solution providers, EDC offers two key benefits:

- A ready-to-use building blocks, forming an extensible foundation for data space components.
- Confidence that their solutions can achieve compliance with DSP and DCP with minimal integration cost.

14.8 Technical Compatibility Kit (TCK)

The Technical Compatibility Kit (TCK) is the backbone of compliance verification. It is a test harness and collection of tools designed to automate the validation of data space implementations

against DSP and DCP. By leveraging shared core libraries, the TCK provides comprehensive tests that cover protocol compliance and interoperability scenarios. Solution providers can run their implementations against the TCK to obtain evidence of compliance. Passing results serve as an objective and transparent proof that a solution adheres to the agreed specifications.

The availability of the TCK ensures that the ecosystem does not fragment into incompatible variants. Instead, it promotes trust and interoperability, which are prerequisites for scaling Data Space adoption across industries and borders.

14.8.1 The Data Space Landscape

14.8.1.1 Data Space Connector Report The Data Space Connector Report is a key regular publication from IDSA offering a comprehensive overview of Data Space Connectors and their role in interoperable data spaces.

In particular, the Data Space Connector Report:

- highlights the importance of Data Space Connectors, explaining what they are and why they are a key element in data spaces.
- it provides a summary of all the key requirements to make Data Space Connectors interoperable (e.g. relying on standards, having clear specifications, enabling semantic interoperability via the Data Catalog Vocabulary (DCAT) and specific vocabularies, etc.) based on the Dataspace Protocol.
- it gives visibility to existing connector implementations, provides details about them and follows their evolution over time.
- it is the reference point for learning and fostering interoperability in data sharing ecosystems.

14.8.1.2 Data Spaces Radar The Data Spaces Radar serves as the central repository for all data space endeavors. It is an accessible tool designed to provide a comprehensive view of various data space initiatives worldwide. Offering insights into the 18 different sectors, global expansion, technical transparency and new stages of development of the data spaces featured in the radar.

14.9 How to contribute

The IDSA Rulebook is published under the CC-BY license. If you wish to contribute, please take a look at our Contribution Guidelines. Please take our Code of Conduct into account.

15 Guiding Principles

16 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.

17 Overarching Considerations

17.1 Overarching considerations of data spaces

17.1.1 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.

17.1.1.1 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 digital 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]. 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. The challenges and objectives are described in the Commission's "2030 Digital Compass"[12]. The Commission proposes several legislative instruments to implement 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 for 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 circumstances.

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, 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 their dynamic evolution over time, while aiming at a common European data space.

Finally, in the "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 protect EU values. One of the priority areas identified is "data standards enhancing data interoperability, data sharing and data reuse in support of the Common European Data Spaces".

17.1.1.2 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 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 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. Consensual agreements in the data space can make this mutually beneficial.

17.1.1.3 S Social The social values embedded in the work of IDSA data spaces are European ideals such as freedom, inviolability, privacy, security, humanity, and respect. Issues such as 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 societal value commitment is *pluralism* of *interoperable* and mutually *respectful* data spaces whose values and priorities are defined in an *inclusive* manner.

17.1.1.4 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 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. Market needs will drive

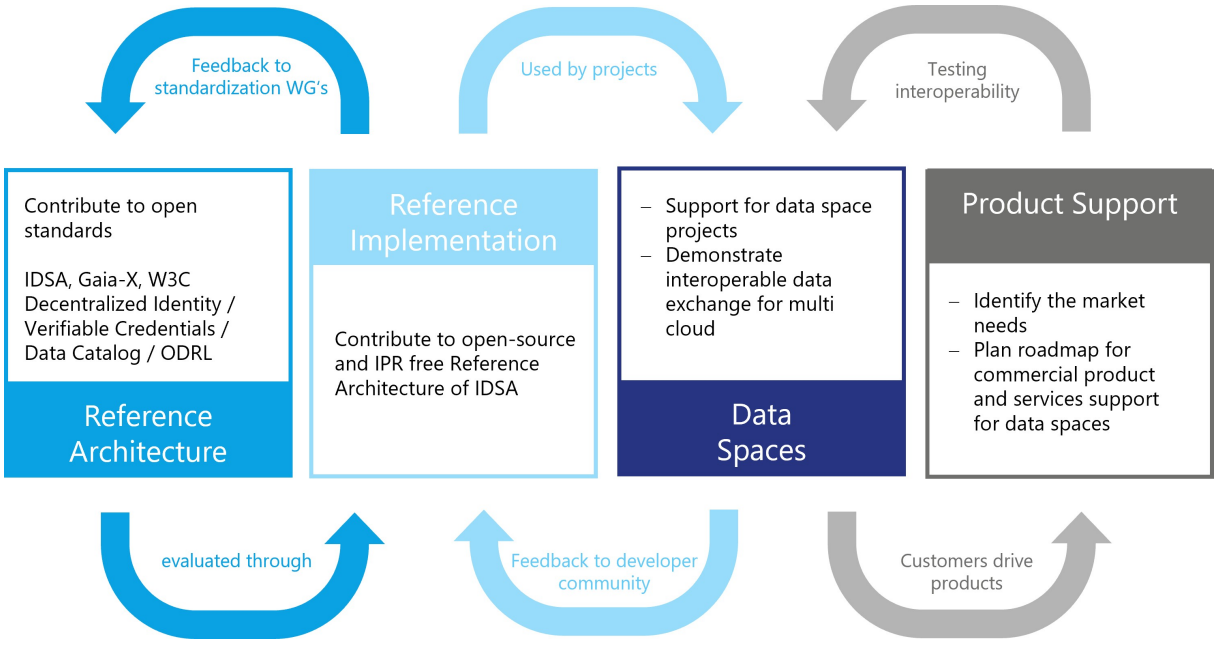

Figure 27: Collaborative Development of Architectures and Implementations in Data Spaces

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 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 further development.

17.1.1.5 L Legal There is a strong connection between political and legal factors. Legislation follows political decisions. Besides knowing the existing legislation, the impact of new and planned regulations based on political developments must be taken into account. Political 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 brings a higher level of regulation to data sharing in the EU, including the Data Governance Act (DGA), the Proposal for Data Act (DA-E), the Digital Markets Act (DMA), the Digital Services Act (DSA) and the AI Act. If a data space operates globally the legal framework becomes more challenging since each country has its own rules and regulations.

17.1.1.6 E Environmental Data usage - collecting, processing, or federation - has a huge and growing impact on our planet. The EU Data Strategy states that making more data available and improving data use is essential to address societal, climate and environmental challenges, contributing to a healthier, more prosperous and sustainable society. It will lead, for example, 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 at 5 to 9% of the global 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 to sustainable digital technologies. AI-based services and state-of-the-art data mining technologies can increase resource efficiency, optimize supply chains, improve coordinate sector coupling and thus lower emissions and add value. Avoiding rebound effects with 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 footprint.

18 Layers of data space governance

18.1 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.

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.

Table: Four Layers to describe data spaces governance

19 Data economy with digital sovereignty

19.1 Data economy with digital sovereignty

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

There are some common rules and guidelines:

- Common definition on lifecycle agreements for IDS-based assets, the IDS standards and services. See appendix “Operational Agreements, Life Cycle”.
- General definitions of necessary processes for development, certification, onboarding, operation and usage. See appendix “Operational agreements. Processes”.

Typical roles 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. A hotline can help with questions (SupportOffice@internationaldatspaces.org).

20 Data space governance framework

20.1 Data Space Governance Framework

The Data Space Governance Framework is the defined set of technical policies, business rules, and regulations that participants in a data space have to adhere to. It is the core agreement between all parties, which defines a data space. The Data Space Governance Authority is mandated to maintain and enforce the Data Space Governance Framework.

The functional requirements section explains how such a framework needs to be translated into a common set of policies.

21 Role models

21.1 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.

21.1.1 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.

21.1.2 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.

21.1.3 Service Provider (intermediary, operator, value-adding services)

In a data space multiple service providers can offer optional capabilities to enable data sharing and data transactions. Some services may be considered as essential, depending on the data space governance framework. Those may be services required to operate a data space, to intermediate in data transactions, or support the value creation process in a data space. Fundamentally, all such service providers are considered to be a participant in a data space and therefore bound to the agreed policies and rules of a given data space.

Operators provide essential services to a data space like Trust Services. Such services are typically mandated by the Data Space Governance Authority and are agreed as service providers for a given data space. Nevertheless, the agreement on such operators may be derived from a certain regualation, that needs to be implemented by a data space governance scheme.

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:

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

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

Such intermediaries may be regulated by local governments like the DGA in the European Union. A detailed analysis can be found in the paper Reflections on the DGA and Data Intermediaries.

Value-added service providers act as a participant in the data space and do therefore stick to the agreements in a data space. Such services aim to enable the value-creation process with a broad set of basic or advanced data processing services. Such optional functionalities are described in the functional requirements section and in the DSSC Blueprint Version 1.

21.1.4 Additional Roles in a Data Space

The DSSC Blueprint Version 1.0 provides a sophisticated anaylsis of additional roles in data spaces, which are not subject of the IDSA Rulebook. Further information on such roles are available in the DSSC glossary. Those roles are grouped into the key analysis areas of the DSSC Blueprint, legal, organizational, business, and technical.

21.1.4.1 Roles from legal defintions Such roles originate from regulations, like GDPR or DGA: * Data Rights Holders are natural or legal persons, holding rights on the data. * Data Recipients are legal or natural persons that act as data consumers and data users. * Data Users are natural or legal persons, which use the data under the given policies and regulations. * Data Subjects are defined in GDPR. * Data Intermediation services or Data Intermediaries are subject of the Data Governance Act. (see Reflections on the DGA and Data Intermediaries)

21.1.4.2 Core Roles

- Data Transaction participant, the term Data Transaction needs further clarification to be integrated in the IDSA Rulebook.
- Data Space Governance Authority is described in detail in the remainder of the IDSA Rulebook.
- Data Rights Holder and their intermediaries are not covered in the IDSA Rulebook.
- Data Provider are covered in the IDSA Rulebook above.
- Data Recipient are covered as Data Consumers in the IDSA Rulebook above.
- Data Space Participant are described in the IDSA Rulebook.

21.1.4.3 Services & service provider roles

- Data space enabling Service, see section above.
- Data space intermediary, see section above.
- Connection-providing Intermediary is not described in the IDSA Rulebook.
- Personal data Intermediary are not described in the IDSA Rulebook.
- Clearing house service providers are described in the IDS RAM and subject of the obverservability capability in the functional requirements section of the IDSA Rulebook.
- Marketplace and other value creation services are covered in the IDSA Rulebook above and in the optional capabilites section of the functional requirements.

22 Layered Approach

23 Understanding Roles and Layers in Data Spaces

Data spaces are multi-layered ecosystems that rely on the seamless integration of technical protocols, business processes, and legal frameworks. One of the foundational challenges in governing data spaces lies in the consistent definition and use of key concepts such as **"roles," "policies,"** and **"contracts."** These terms often carry different meanings across domains. This chapter establishes a clear separation between technical and non-technical interpretations of such concepts to support the development of interoperable and trustworthy data spaces.

23.1 Layers of a Data Space

Data spaces can be structured into three primary layers, each serving distinct functions:

1. **Technical Layer** – Encompasses the architecture and protocols (e.g., Dataspace Protocol (DSP), Decentralized Claims Protocol (DCP)) that facilitate secure and interoperable data exchanges.
2. **Economic Layer** – Manages the services, interactions, and workflows that enable value generation and marketplace activity. Notably, terms for this layer are also Business or Operational Layer.
3. **Legal and Governance Layer** – Enforces rights, obligations, and regulatory compliance across participants.

Figure 28: Layers of Data Spaces

These layers interact but must be conceptually separated to ensure clarity and reduce ambiguity in roles and responsibilities.

23.2 Clarifying the Concept of Roles

The term **"role"** is context-dependent and must be clearly scoped:

- At the **technical level**, there is only one fundamental role: **participant**. A participant acts as a **data provider**, a **data consumer**, or both within the data exchange protocol.

- At the **business level**, participants may take roles such as **data intermediary**, **marketplace operator**, **auditor**, or **service provider**.
- These business roles do not exist independently at the technical layer but are mapped onto the core participant role based on the services performed.

Maintaining this distinction ensures that governance models remain technically sound while accommodating diverse business scenarios.

23.3 Distinguishing Data Spaces from Trusted Data Transactions

A clear differentiation must be made between **data spaces** and **trusted data transactions (TDTs)**:

- **Data spaces** are decentralized infrastructures that enable sovereign data exchange based on open standards. They preserve participant autonomy and operate without mandatory intermediaries.
- **Trusted data transactions**, as under current standardization in the European Commission's Standardization Request on a Trusted Data Framework in CEN/CENELEC JTC 25, can also be associated with the EU *Data Governance Act*. They can also be related to **data intermediaries** and **service orchestration**. Such models prioritize regulatory alignment and controlled environments.

While trusted data transactions may operate within data spaces, they are conceptually distinct. Equating them risks narrowing the scope of data space implementations and excluding more decentralized or peer-to-peer configurations.

23.4 Participation and Representation

Participants in a data space are defined by their ability to exchange data via technical protocols. This has several implications:

- **Organizations**, not individuals, are considered technical participants. These organizations are represented by **software agents** capable of executing data space protocols.
- **Natural persons** interact with data spaces indirectly through applications or services operated by organizations.
- **Trust anchors**, **identity providers**, and **regulators** may influence data transactions but do not participate directly unless they act through technical interfaces governed by data space rules.

This model preserves the integrity of technical interactions while allowing flexibility in higher layers.

23.5 The Role of External Actors

Entities that provide static resources—such as ontologies, schemas, or public credentials—may support the data space but are not considered participants unless they actively engage via governed interfaces. For example:

- A web service that hosts a data sharing ontology is not a participant but serves as an **external reference**.
- A trust framework provider may act as a **participant** if it delivers services subject to data space governance policies.

Participation requires **governance commitment** and **technical integration**.

23.6 Implications for the Rulebook

The Rulebook should reflect these principles clearly:

- The only **technical role** is the **participant**, which may act as data provider, consumer, or both.
- **Business roles** are supplementary and must be defined within the business or legal governance layers.
- **Visual representations** of data space structures must be clearly labeled to indicate whether they depict technical, business, or legal perspectives.

Such clarity supports interoperability, ensures accurate alignment with regulatory frameworks, and promotes broad adoption across sectors.

23.7 Overview on roles and Layers

In line with the description of the role models and the layered approach, the diagram below presents an overview on roles in data spaces and their affilation to the layers.

This diagram is a foundation to depict the typical use cases of the roles in relation to data spaces.

23.8 Conclusion

Effective data space governance depends on the precise use of terminology and clear separation of concerns across layers. Establishing the **participant** as the core technical role, while accommodating richer business and regulatory interactions above it, ensures a scalable and interoperable foundation. This layered perspective will guide the elaboration of rules, responsibilities, and interactions in subsequent chapters of the Rulebook.

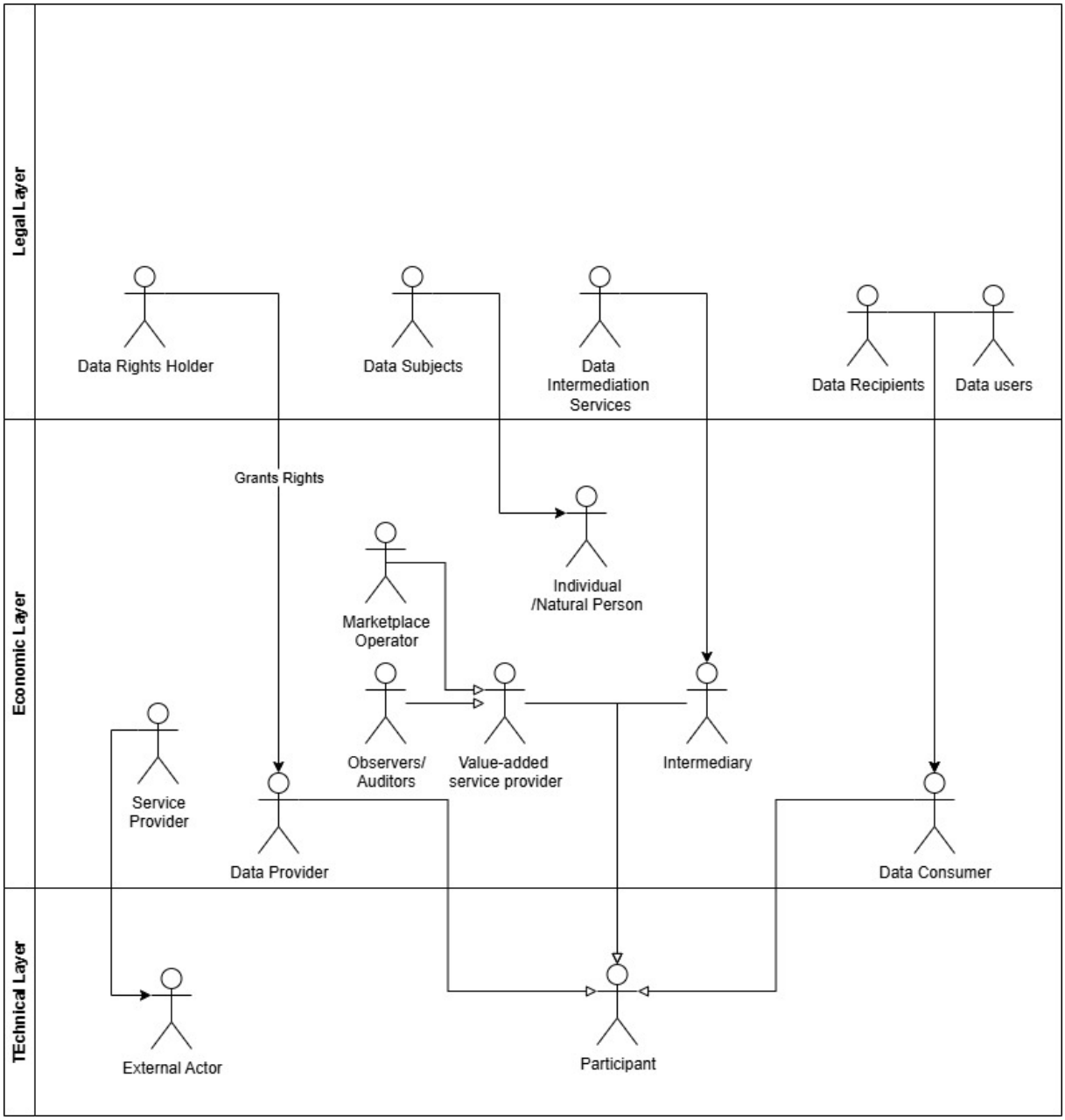

Figure 29: Overview on roles and their affiliation to layers in data spaces

24 Functional Requirements

25 Functional requirements for a data space

This section of the Rulebook 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 enable control over data sharing and create value for all involved parties.

A data space is both a multi-organizational agreement and a supporting technical infrastructure 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 of data consists of peer-to-peer interactions. 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 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.

> The functional requirements section refers to all involved roles as ‘participant’ in a data space. This underlines the need for all parties involved in a data space and in the exchange, sharing, and usage of data to adhere to a common set of rules, the policies provided by the data provider, the rights granted by a data rights holder, and given regulations.
>
> An overview on roles in a data space is given in section 2 on guiding principles

26 Achieving digital sovereignty

26.1 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 assets.

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 for necessary functionality for all.

The data space governance authority (DSGA) 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 data space governance 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 such rule set is the book of law for the membership. When a data space operates in a regulated industry, there are laws and regulations 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.

27 Foundational concepts

27.1 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

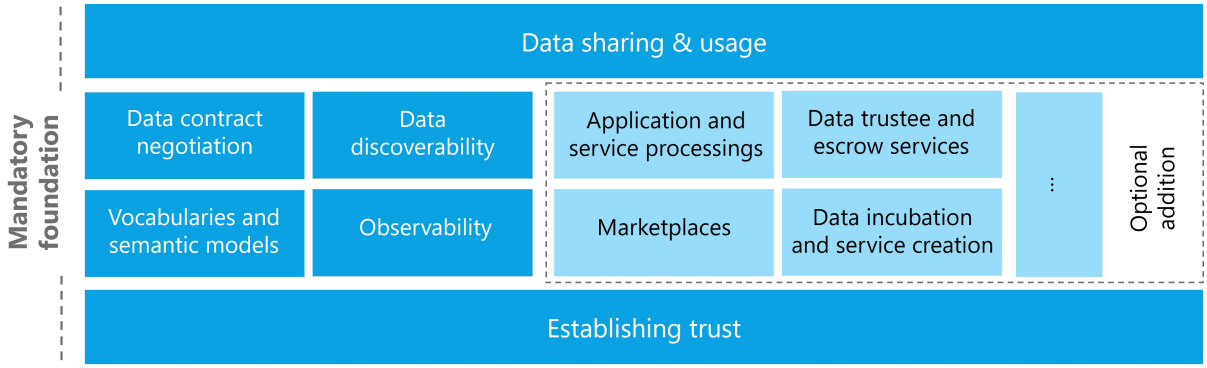

Figure 30: Foundational Concepts in data spaces

28 Establishing trust

28.0.1 Establishing trust

Establishing 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. #### Attributes & self-descriptions

When people build trust with 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, these attributes are matched against (personal) policies. If a sufficient number of policies are met, 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 with the requirements, policies and rules of the data space, the participants, and individual data contracts.

A data space needs to define policies that specify what attributes an applicant must meet 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 governance authority (DSGA). 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 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 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, these services 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 similar to mechanisms that citizens use in their daily lives: 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 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 must follow the laws of the country it is based in to obtain a valid company registry ID issued by its government.

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 DSGA 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 use of the concepts described above, the DSSD needs to contain information about which trust anchors and trust frameworks are accepted as roots of trust. Is it a sovereign entity 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, a potential participant can make the decision whether to

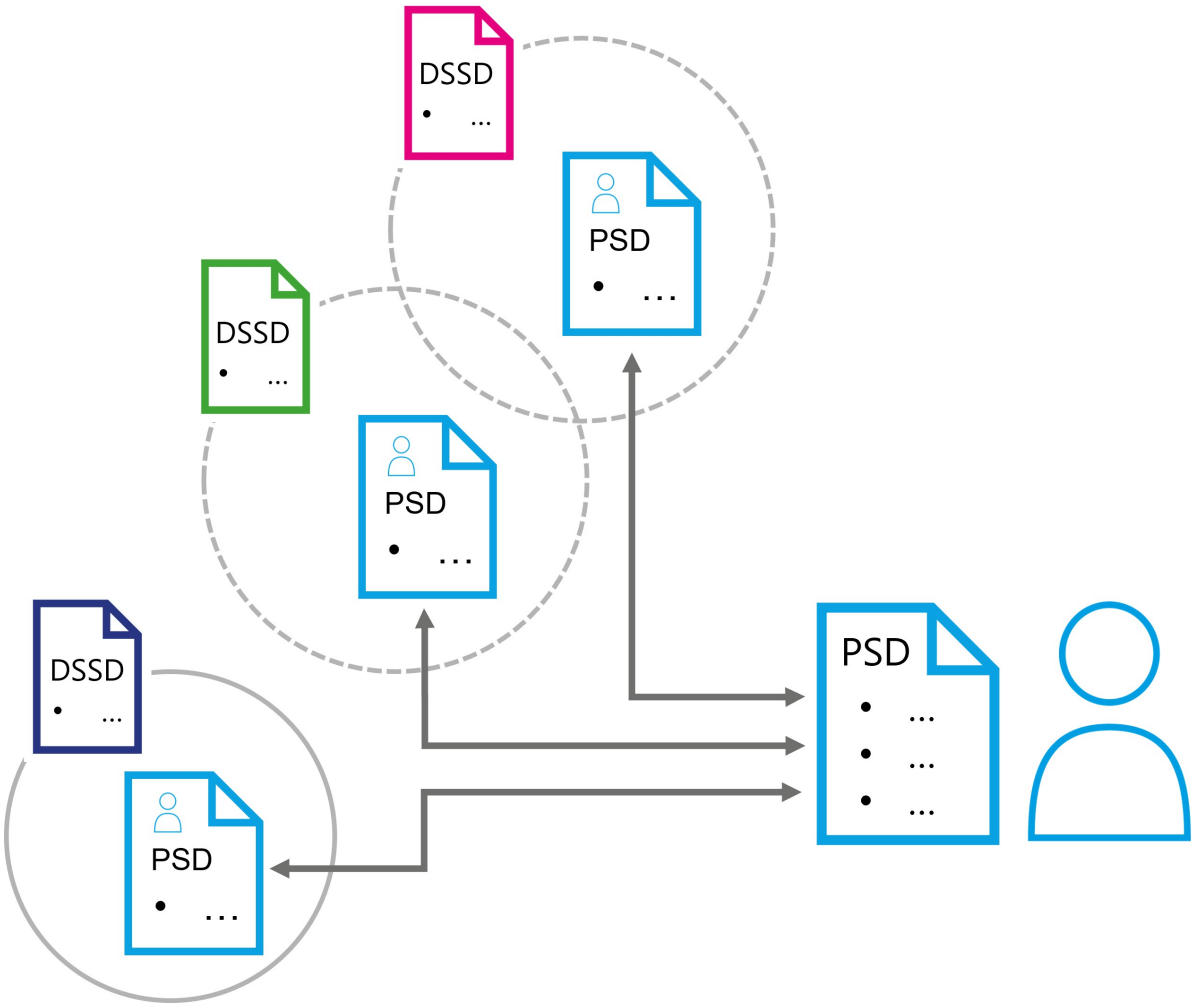

Figure 31: Self Descriptions in data spaces

trust the data space and its members or not.

The DSGA is also responsible for issuing membership credentials. It ensures that an appropriate mechanism is provided for identifying and verifying membership. In a centralized data space this could be the issuance of a data space specific identity to interact with other members. In a largely decentralized architecture, it could be the issuance of a tamper-proof credential, such as a W3C verifiable credential (VC) which provides proof of the attribute of membership.

The DSGA also performs other functional roles not directly related to building trust but necessary for the operation of a data space. These are 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 DSGA is required, or a more federated or even fully decentralized model is appropriate must be reasoned over when the data space is founded, as these architectural choices are 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.

28.0.1.1 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 essential to understand.

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 health 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 communicate them clearly. Participants may always choose additional policies in their data contracts to further restrict access and use.

In a centrally managed data space, the DSGA might simply define the ontology of policies. 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 restrictions: 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 to the same industry association, allow to process data under the condition only anonymized results are produced, and then 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 can verifiably prove, can join. These could be policies that verify the 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 DSGA 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

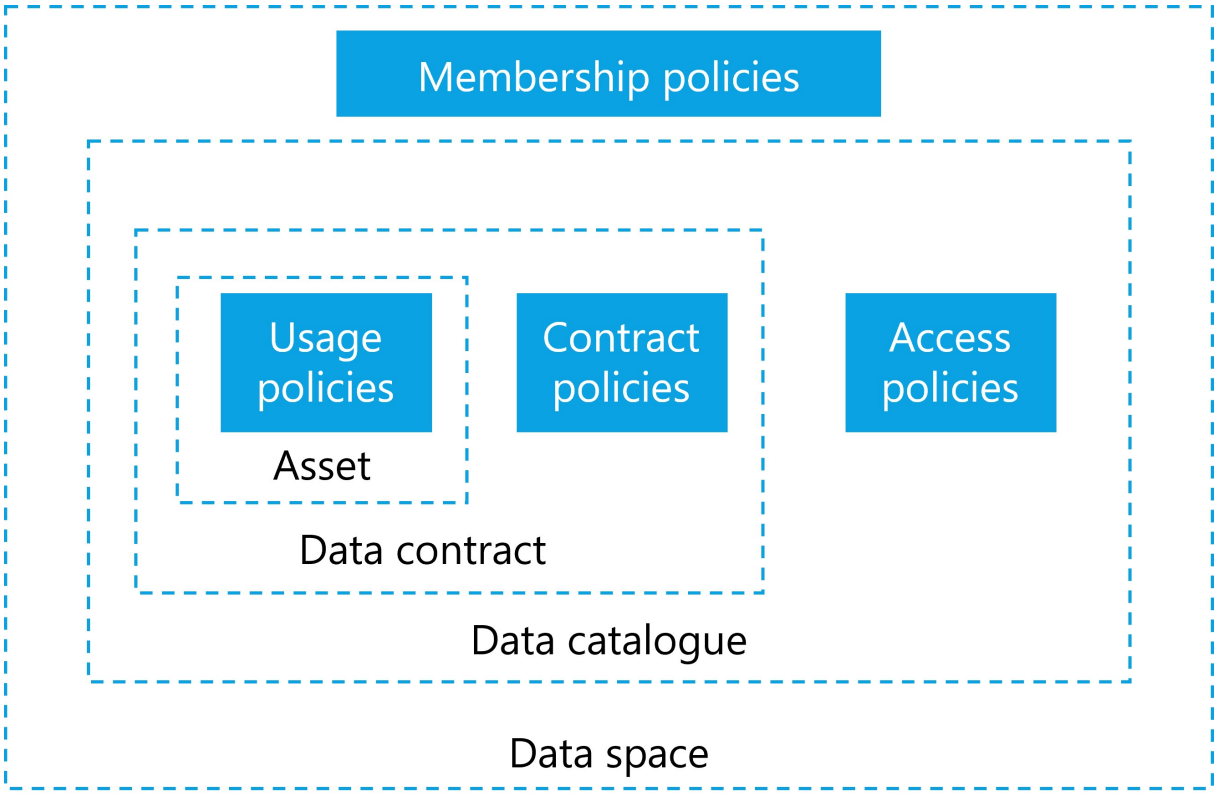

Figure 32: Different policies in data spaces

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 matching access policies and participant attributes. In a scenario where contract offers should be made visible to 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 non-members (in case the catalog endpoint is publicly accessible).

Each participant can define such policies, whether providing or consuming data. For example, a participant interested in 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. 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 review attributes that must 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 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 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.

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 low importance or data not under a specific legal protection, it might be too expensive to build a system that guarantees control - it may be sufficient to simply monitor 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 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	**ProtectionNeed**	**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.

28.0.1.2 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 a set of multiple atomic expressions (e.g., the other entity is under a specific jurisdiction and the 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).

Many situations will required attributes that are complex and might require complex workflows that 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.

28.0.1.3 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 DSGA 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.

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 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 the membership

policies 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. 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 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 implemented and expressed through the functional role of the data space governance authority is highly dependent on the needs of the data space and is the most important decision to be made when designing a data space.

28.0.1.4 Participant information Information about a participant 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 DSGA and may be a part of the membership policies for the data space. In many cases, the format and ontology of the PSD also depend on the selected trust anchors and trust framework. For example, 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 Gaia-X self-description attribute definitions. A data space might require multiple self-description ontologies (e.g., one trust anchor specific and one industry specific) which can lead to ambiguity or conflict of definitions, which have to be resolved by the DSGA.

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 must 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.

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.

29 Data space participation

29.0.1 Data space participation

Participation in a data space is based on fulfilling all the policies, rules and procedures that are mandatory for membership. In its simplest form, these may just be technical or automatically verifiable policies. In more advanced cases, these can be more complex policies and rules that potentially require lengthy workflows with human interaction to verify eligibility to join 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 (details can vary due to the design and purpose of the data space):

1. Candidate discovers the data space and the corresponding DSSD
   This can be achieved through human interaction, a website of the data space, finding the DID of the data space in some registry or through automated discovery protocol of existing participants among other things.
2. 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.
3. Candidate evaluates the policies and rules and prepares additional information needed for the requirements when applying for membership in the data space.
4. When all information and necessary proofs are collected the candidate applies for membership through the registry function of the DSGA. The technical implementation of the data space registry might vary based on the requirements.
5. The DSGA requests proofs for all policies. This might include VCs and proof of technical capabilities, but also workflows including human interaction (e.g., signing a membership contract).
6. Once all policies have been satisfactorily processed the DSGA issues a VC/ proof of membership and sends it to the candidate, moving them from applicant to participant.
7. The new participant sets up all the necessary technical components for participation in the data space.
8. 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).

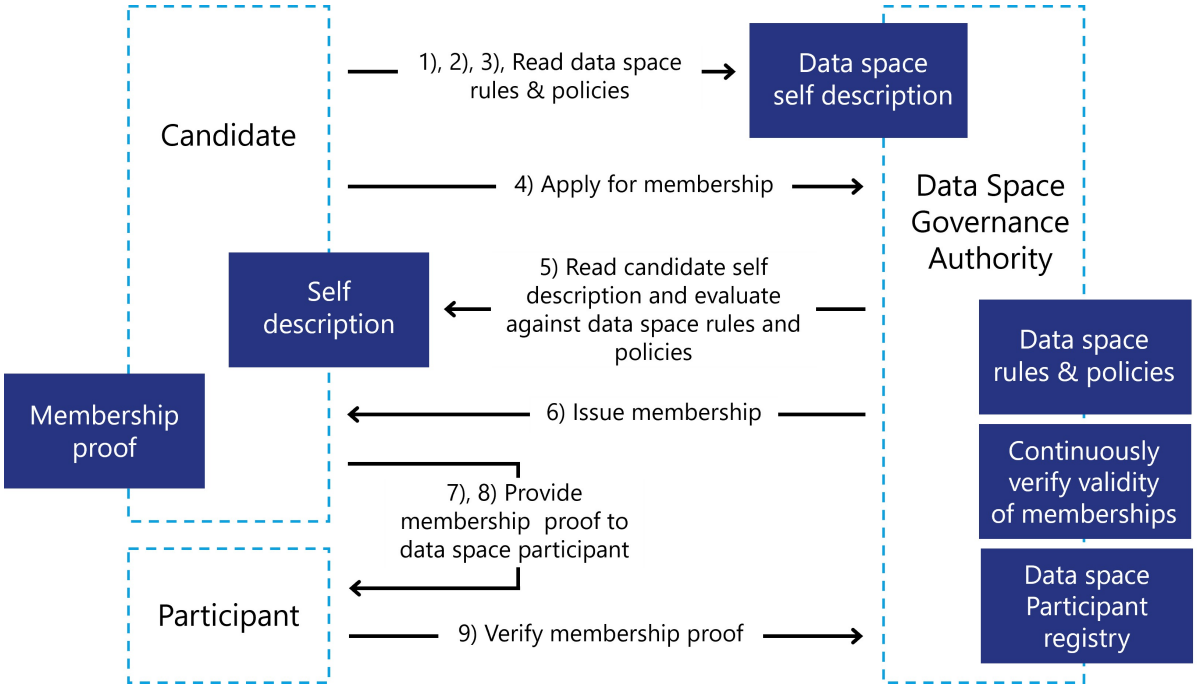

Figure 33: Onboarding in data spaces

30 Creating a data space

30.0.1 Creating a data space

After discussing how to join a data space the question is: How do you create a data space? The answer depends again on the purpose of your data space and the needs of its participants. Regardless 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 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, or a large community with many participants. Some data spaces are created for a narrow use case and purpose others for many use cases that are relevant for a group of participants.

Many decisions need to be made when designing the data space, here some of the more common ones:

- Is the membership closed to a small, known group or open to a larger range of participants?
- Do you want 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 expected from the participants?
- What type of data is shared and for what purpose?

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

Once all design decisions are made, the functional elements are planned:

- Rules: What behavior and skills (technical and organizational) are required?

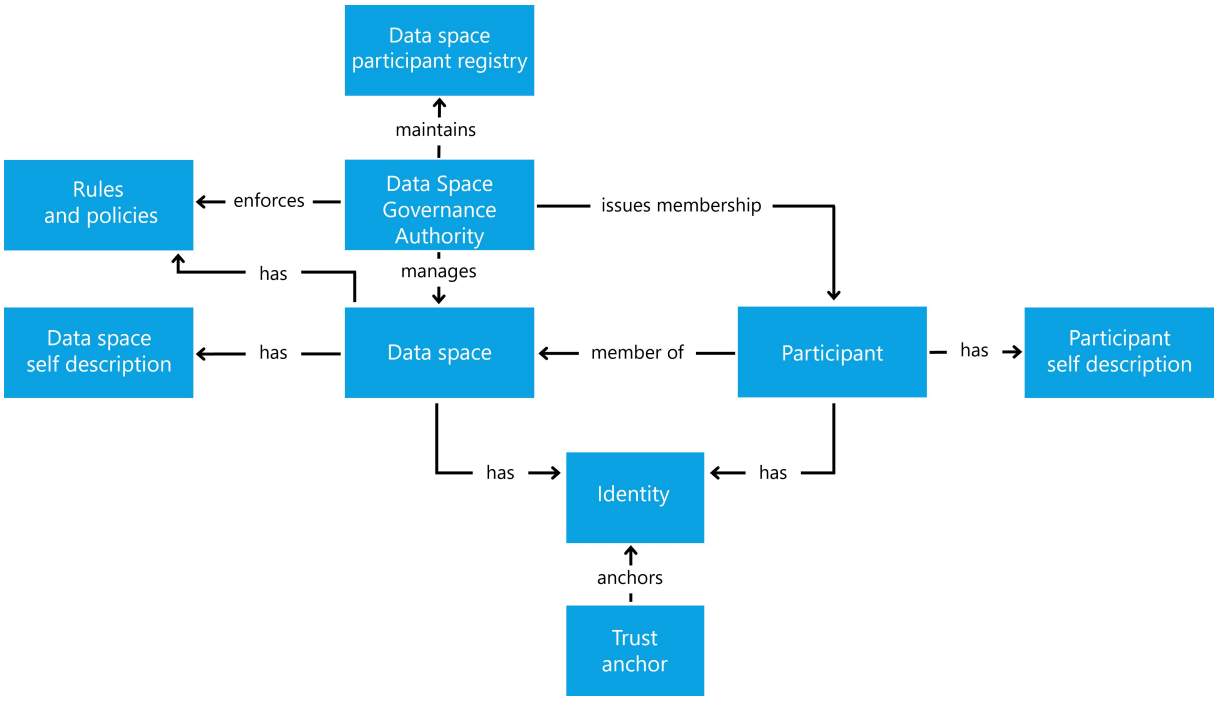

Figure 34: Overview of Data Space entities

- 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 clarify the architecture pattern for the data space, which will also mandate a specific design of the data space governance authority. Now the DSGA 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
  - 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 DSGA 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 provides. Additional functions and services such as marketplaces, data escrow services, processing services and applications might be provided as optional elements.

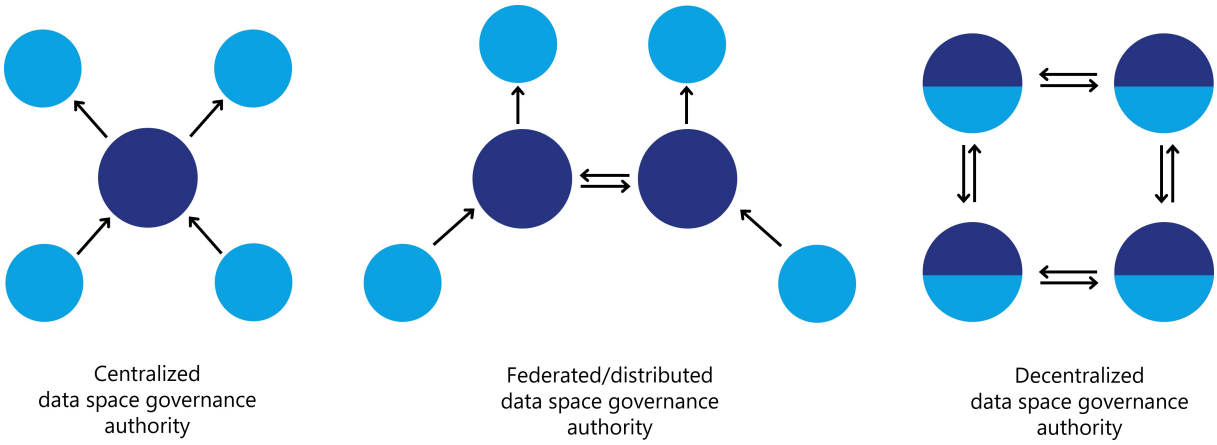

Figure 35: Variants for data space governance authorities

31 Data discovery

31.0.1 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.

32 Catalog(s)

32.0.1 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 the data is open or protected. Information about the data needs to be published with an agreed-upon vocabulary for querying and with controls that regulate access to the catalog items.

Two participants can share data directly communicating off- or online without the need for a catalog. But for more participants 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 must allow federated searches of data assets in catalogs at multiple sites.

Catalogs don't provide the data asset itself, but they provide 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, 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 DSGA, to evaluate their capabilities:

Catalog architecture	Advantages	Disadvantages
Centralized catalog	No deployment by individual participants	A central gatekeeper can arbitrarily exclude participants and their data from the catalog
	Central control – a gatekeeper can regulate which entries are permissible and which are not	Single point of failure
	Easy discovery as only one catalog needs to be queried	Potential performance bottle neck
		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	Additional replication mechanisms are needed
	Federated control – voting mechanisms for content control can be implemented	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	Every participant needs to run a catalog component
	No interference in the interaction between two participants through a 3rd party	A list of available catalogs needs to be either centrally provided through the DSGA or discoverable through a peer-to-peer protocol

Catalog architecture	Advantages	Disadvantages
	Data Space as a whole is more resilient towards cyberattacks even though individual members can experience outages Easier to scale	Participants need to crawl each other's catalogs to see which items are available

32.0.1.1 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 that a participant is allowed to see that the data exists. The permission to access is part of the data contract negotiation. Any catalog must implement attribute-based access control (ABAC) through access policies.

The most common access filter is that a participant 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 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.

33 Data sharing

33.0.1 Data sharing

Once a participant has joined a data space and discovered available data contract offers, the mechanism of data sharing is initiated. Data sharing 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 ranges 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 the source and algorithms and processing code are copied to the data location for in-place processing. Data Sharing does not 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.

33.0.1.1 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 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 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 is responsible for sharing the data. Or there might be data assets that can't be immediately shared after the agreement is reached (e.g., an event notification that can only be consumed until 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 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

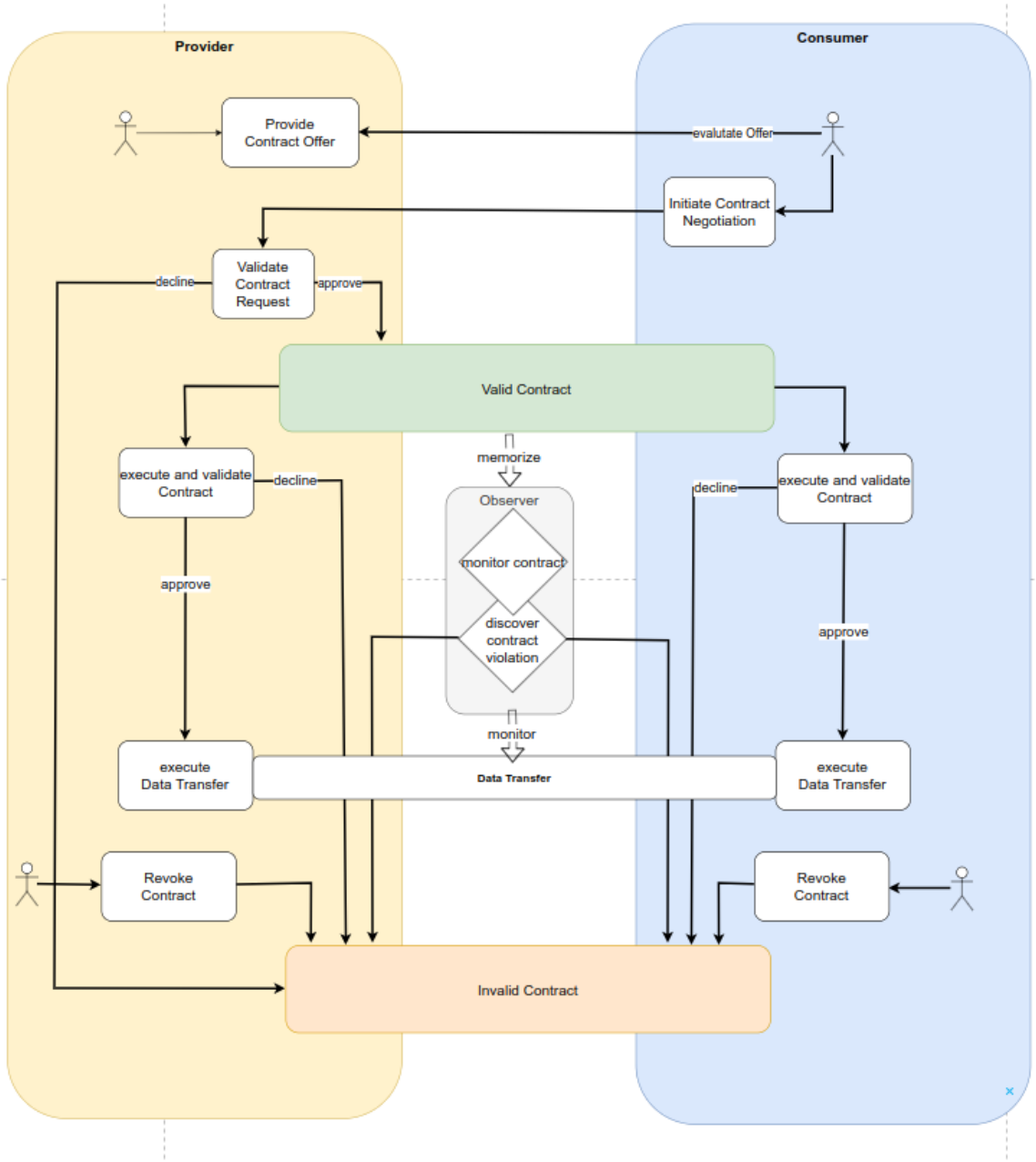

Figure 36: Data sharing contract negotiation

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 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 in 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.

34 Observability

34.0.1 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 two shortcomings when implementing large-scale data spaces: It presents an additional vulnerability that could affect the sharing of mission critical data. And 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.

To address these risks, having at least a federated model of observers is recommended to distribute the information, load, and potential for error. 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. 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 limit access to sensitive observation data to observers that are participants of the data space, have special credentials which qualify them as trusted auditors and 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 DSGA 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: a payment clearance service, notary services, regulatory reporting, and the like.

34.0.2 Vocabularies

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
  For example, if a data space wants to restrict membership to companies with a HQ in certain countries. 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 be 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)

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 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.

34.0.3 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 DSGA to establish trust.

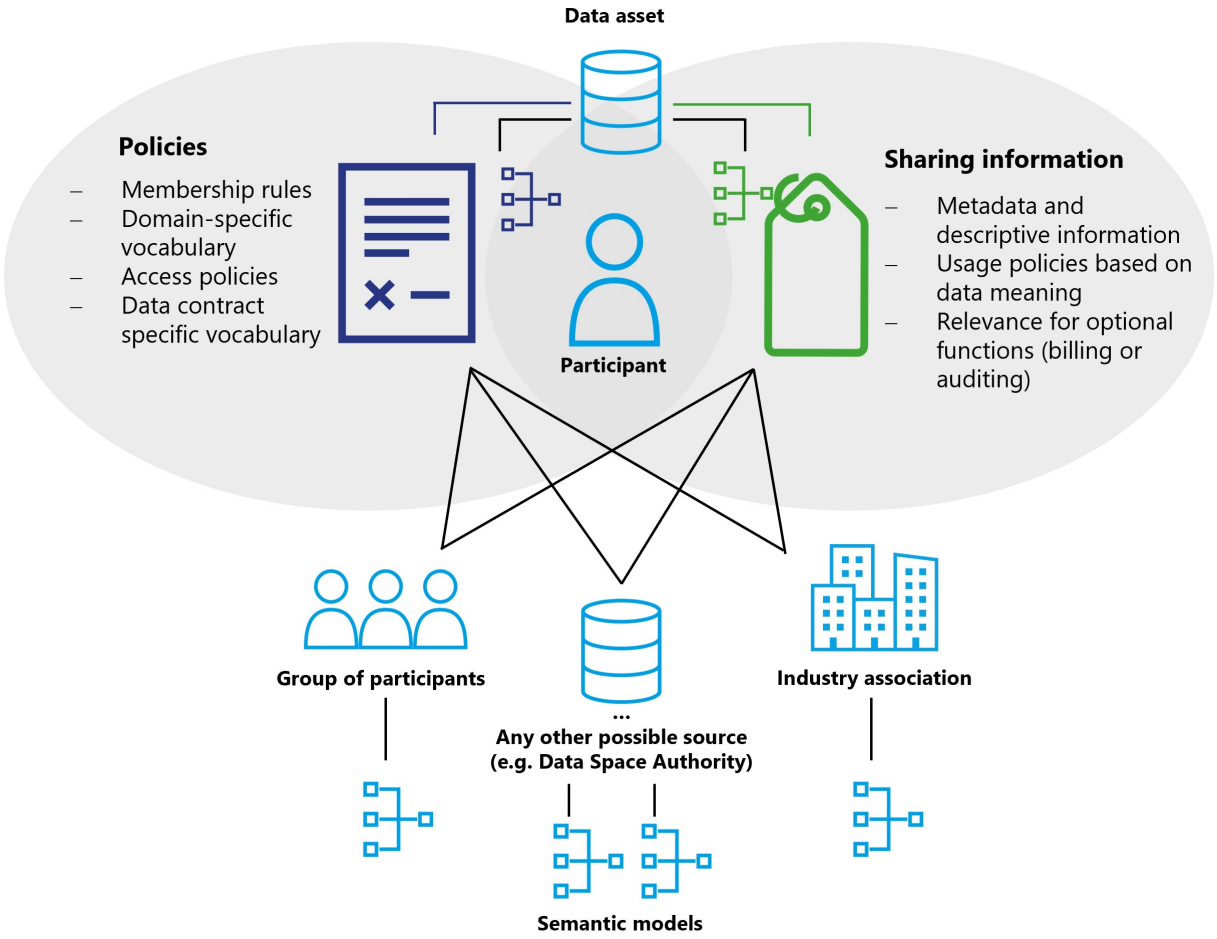

Figure 37: Vocabularies and their relationship to data assets

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

In general such optional functions can be distinguished as intermediary functions or value-creating functions. Intermediaries can participate in data spaces as value-creating services or functions.

Intermediaries are considered as optional in data spaces. Due to certain regulations like the Data Governance Act, such intermediaries may require additional governance.

Value-adding services may be realized by intermediaries or as function of a data space participant. Such value-adding services are not subject to the IDSA Rulebook, but are explained in the DSSC Blueprint Version 1 in more detail. The IDSA Rulebook provides a limited explanation below.

34.0.3.1 Marketplaces 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 many dataspaces require not only searching for 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. Similar to how resources can be bought and sold on exchanges, functions 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.

34.0.3.2 Processing services A data space can have participants that do not offer their data and are not the end users 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 these participants might offer their own infrastructure as part of the contract and use policies to control the use of their resources.

Many data spaces can be built on top of the peer-to-peer model, such as a data supply chain where data assets pass through multiple processors before reaching the end user. The implementation and capability of these services again depends on the architecture, policies, and rules of the data space.

34.0.3.3 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 to provide trusted services and not to use the data.

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

34.1 Technical components of a data space

34.1.1 Data space governance authority services

Several services are required that represent the functional role of the data space governance authority (DSGA) 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.

34.1.2 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 manage the identities for all participants, every other service depends on this central verification, and decentralized designs are no longer fully feasible.

Which mechanism to use 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	Simple management for DSGA	Low autonomy and sovereignty of participants
	High degree of control for DSGA	Single point of failure
	Traditional, well-known technology stack	Single point of attack
		Harder to manage for participants
Decentralized identities	Full autonomy and overeignty for participants	Complexity: DSGA management requires decentralized protocols
	Low resourcing need for DSGA	Lower degree of control for DSGA
	Easy to manage for participants	New and partially unfamiliar technology stack
	Harder to attack	

34.1.3 Catalog

The catalog component supports the search for available data contracts. Information about data contracts can be exchanged between participants without the use of a catalog by 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 governance 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 DSGA 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.

34.1.3.1 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.

34.1.4 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.

34.1.5 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.

35 Vocabularies

35.0.1 Vocabulary

The semantic model for the policies and self-descriptions required to join the data space is provided by the DSGA. 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 DSGA must decide how semantic models are provided, whether by reference to a known, standardized schema externally or through a vocabulary service provided by the DSGA or specific participants.

Individual participants may 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 DSGA 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.

35.0.2 "Central," or "federated/distributed," or "decentralized"

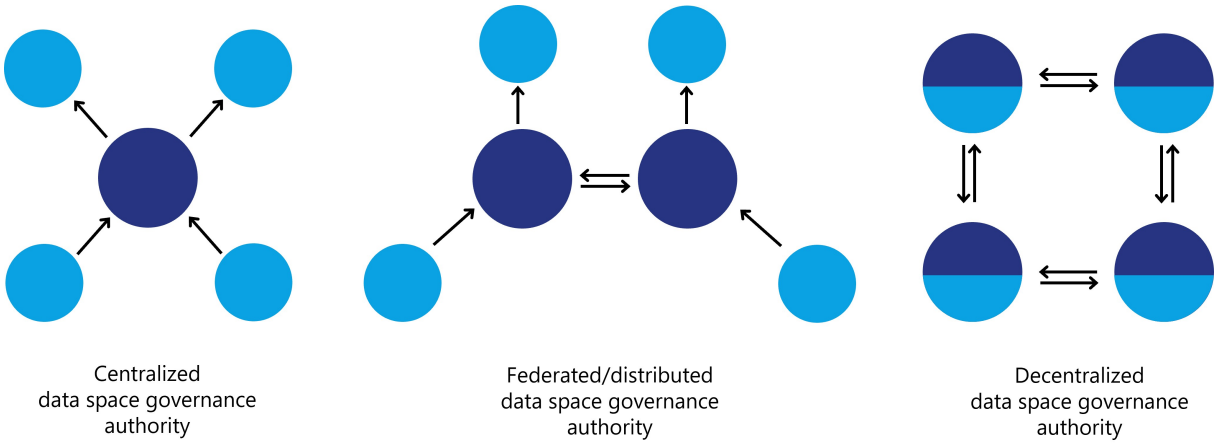

Figure 38: Variants for Data Space Governance Authorities

35.0.2.1 Centralized data space governance authority In a centralized DSGA 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 audits.

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 known location to discover

available data and queries only need to be made at one endpoint and data contract offers are returned from multiple participants. But it poses the risk that the entity controlling the catalog 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 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. An infiltrated central identity provider or catalog could create more damage than if a 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 not be a problem.

35.0.2.2 Federated / distributed data space governance 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 supports the federation functionality of a data space and serves a functional role such as identity or catalog.
- Federated service describes the implementation of any service as a distributed service in a data space, including but not limited to any of the federation services.

To maximize the sovereignty and autonomy of participants in a data space, every participant must be free to act 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, 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.

35.0.2.3 Decentralized data space governance 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 DSGA, 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 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 DSGA to the participants. For example, in a centralized model, the DSGA 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 about their available assets.

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-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.

35.0.3 Decision areas

35.0.3.1 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 move along a spectrum. The goal is to increase sovereignty and autonomy until a desired threshold is reached. In that sense, the concept is similar to that of privacy.

35.0.3.2 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.

35.0.3.3 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.

35.0.3.4 Control In this context, a high level of control means that the entity operating the DSGA 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.

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

35.0.3.6 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 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.

35.0.4 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:

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 DSGA services, and participants can simply query one catalog and rely on the DSGA 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.

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.

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.

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.

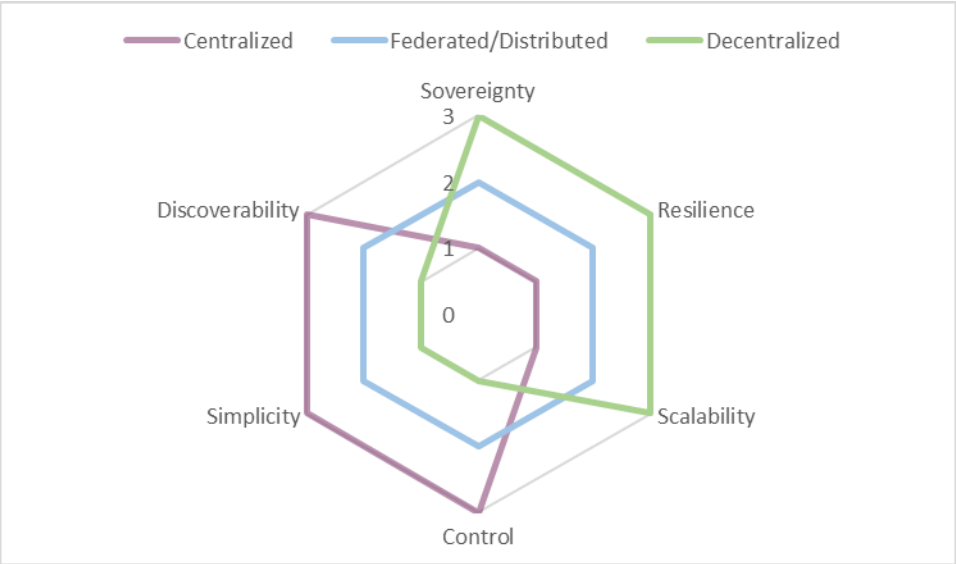

Figure 39: Comparison of models for decision support

36 Interoperability in Data Spaces

37 Data Spaces Interoperability - How to achieve Interoperability within a Data Space and across multiple Data Spaces

37.1 Motivation for interoperability

Data is one of the most valuable assets in the digital economy, but its potential value can only be realized if it can move and interact with other data to produce insights that create value. For this, it must be possible for data to be shared and reused in a trusted way. Interoperability is the ability of different systems and organizations to exchange, understanding, and use of data, is essential for enabling data sharing and creating value in data ecosystems. Data Spaces help to establish a common understanding of trust, and provide a mechanism to establish sharing contracts, which include access and usage policies that ensure the protection and accountability of data providers and data consumers. As Data Spaces become more prevalent and diverse, there is also a growing need for intra- and cross-Data Spaces interoperability.

Different Data Spaces may have different goals, architectures, business models, and governance structures, depending on the authority or community that drives them. To avoid fragmentation and duplication of efforts, participants in these Data Spaces need to communicate in an interoperable way with each other and across multiple Data Spaces, following common standards and principles.

Interoperability can be achieved at different levels, depending on the degree of integration and alignment of the data and systems involved. Two well-known frameworks that define interoperability levels are the ISO/IEC 19941 standard for cloud computing interoperability and portability, and the European Interoperability Framework for public services. Both frameworks identify four main levels of interoperability: technical (transport & syntactic), semantic, organizational, and legal:

- Technical interoperability refers to the physical and logical connections between systems and data sources, such as protocols, interfaces, and formats. This includes syntactic interoperability which refers to the structure and syntax of the data exchanged, such as schemas, models, and vocabularies.

- Semantic interoperability refers to the meaning and interpretation of the data, such as concepts, relationships, and ontologies.
- Organizational interoperability refers to the processes, policies, and governance of data sharing, such as roles, responsibilities, and agreements.
- Legal interoperability refers to the acceptance of legal equivalence of contracts and contractual clauses between different data ecosystems. These ecosystems can have differences on multiple dimensions, based for example on industry regulations, or national laws but also contractual statements with identical wordings might have diverging interpretations in different data ecosystems.

In this chapter, we discuss the challenges and opportunities of achieving cross- and intra- Data Space interoperability at these levels, and propose a roadmap for developing a common framework and best practices for Data Spaces.

37.2 Guiding principles for Data Spaces

As described in the previous sections, there are guiding principles in Data Spaces that are not only shaping the functional requirements, but also take effect in reasoning over interoperability. Those guiding principles are the foundation of any interoperability framework for Data Spaces. Let's have a closer look at those fundamental principles again:

1. Self-determined control of data use (Data Sovereignty as also part of ISO/IEC CD TS 10866, Framework and concepts for organizational autonomy and digital sovereignty) is of utmost importance and should be the ideal vision that each Data Space thrives to enable.
    1. Participants have autonomy and are able to act with choice
    2. Participants have agency over their data assets
2. A Data Space creates a context of trust
3. The Data Space Governance Authority is a governance body for a Data Space

To better understand the model above we need to understand Data Spaces differently at different layers. There is a legal layer where a Data Space is governed by legal contracts to join a consortium that is responsible for the Data Space. This can be a not-for-profit organization where participants join as members to jointly agree on what the rules of the Data Space are, but also can be driven by a single entity that dictates the rules of the Data Space. Both models and everything in between is possible and tradeoffs need to be reasoned over and decisions made when the legal layer of the Data Space is being defined. There can even be Data Spaces without any organization at the legal layer, purely governed by measures provided through the Data Space Governance Authority (DSGA).

The DSGA is a logical function in the Data Space and while it will be quite common to combine the DSGA with the legal organization of the Data Space, it is also possible that a DSGA exists without any legal organization operating it. E.g. a DSGA could be just a set of policies passed around between Data Space participants without any single owner, just been agreed on by a consensus algorithm between participants.

The DSGA is also responsible for the semantic models of the Data Space and thus has a huge influence on the interoperability at that layer.

Taking the guiding principles above into account leads us to the conclusion that interoperability is a shared responsibility between the participants and the DSGA.

With more autonomy and agency, a participant can act with the more responsibility for ensuring interoperability layers with the participant. With less autonomy and agency, more interoperability responsibility moves to the DSGA and legal organization layers thus lessening the burden of interoperability on the participant.

Therefore, it is fair to say that more autonomy and agency of participants also comes with increased responsibilities for the participants.

37.3 Interoperability Models

When talking about interoperability in Data Spaces we need to separate the discussion between two main interoperability models:

1. the interoperability within a Data Space between individual participants (and also with the DSGA of that Data Space), and
2. the cross-Data Space interoperability where a participant wants to access data from two different Data Spaces.

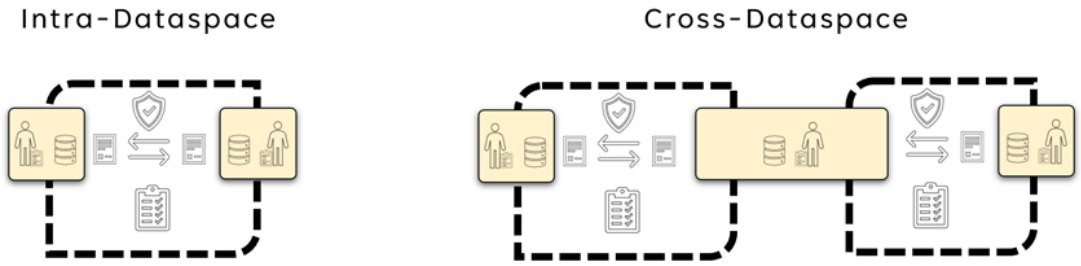

Figure 40: Interoperability Models

Intra Data Space interoperability is about the interoperability within a Data Space. This focuses on how participants interact with each other, and as well as with the DSGA. The DSGA defines what rules govern the Data Space. This includes also which version of the Data Space Protocol needs to be used, what identity protocols and standards to use, which Trust Frameworks are accepted, what semantic models need to be understood, and so on. Participants have the responsibility to at least support and understand the protocols and models that the DSGA mandates but can also support additional versions and semantic models.

Cross-Data Space Interoperability refers to the interoperability required for one entity to participate in two Data Spaces. As it is the participant that wants to access data from two different Data Spaces most of the responsibility for interoperability falls on the participant. First of all, the participant needs to become a member of both Data Spaces, thus fulfilling the membership rules to be able to join both Data Spaces. This implies that the participant is able to support all the protocols and semantic models that both Data Spaces require. Should those not be identical it is up to the participant to be able to support the right protocols and their version in each Data Space and potentially do any necessary mappings. Another option is when the DSGA, as well as the legal entity operating the Data Space (if such exists) can support participants by agreeing with other DSGAs and legal entities from other Data Spaces on supported protocols and semantic models. This can greatly reduce the burden on the participants in sharing data with and using data from multiple Data Spaces.

37.4 Interoperability Standards

There are two noteworthy standards when it comes to interoperability, first the ISO/IEC 19941 – Cloud Computing Interoperability and Portability and second the European Interoperability Framework. The Regulation of the European Parliament and of the Council on harmonized rules on fair access to and use of data (Data Act) references both standards in its provisions for interoperability.

Chapter VIII of the Data Act provides for essential requirements to be complied with regarding interoperability for operators of data spaces and data processing service providers as well as essential requirements for smart contracts. The chapter also enables open interoperability specifications and European standards for the interoperability of data processing services to promote a seamless multi-vendor cloud environment.

Let's investigate the facets of interoperability as defined in those standards a bit closer.

First the ISO 19941 Interoperability facets:

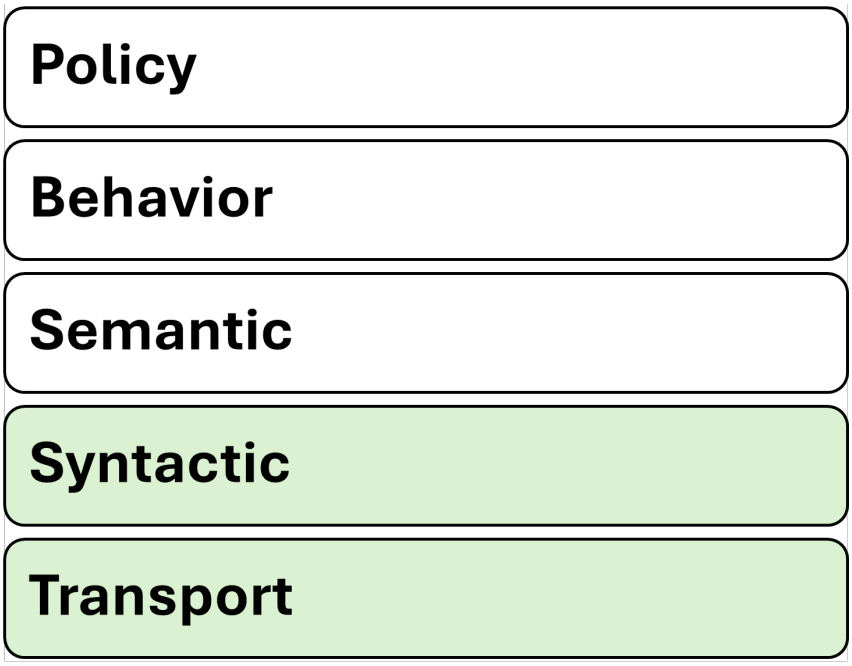

Figure 41: ISO 19941 - Cloud Computing Interoperability and Portability

And second the European Interoperability Framework facets:

Note, that while the EIF has only four layers, it is clearly visible that with the five layers of ISO/IEC 19941 the technical layer is split into two sub-layers: the transport and the syntax.

37.5 Interoperability facets in Data Spaces

Let's investigate the 4 facets of interoperability and how they can be applied to Data Spaces.

37.5.1 Technical

The basis for technical interoperability in Data Spaces is the Dataspace Protocol(DSP). This protocol provides a set of specifications designed to facilitate interoperable data sharing between entities governed by usage control and is based on web technologies. These specifications define the schemas and protocols required for entities to publish data, negotiate agreements, and access data as part of a federation of technical systems that form a Data Space.

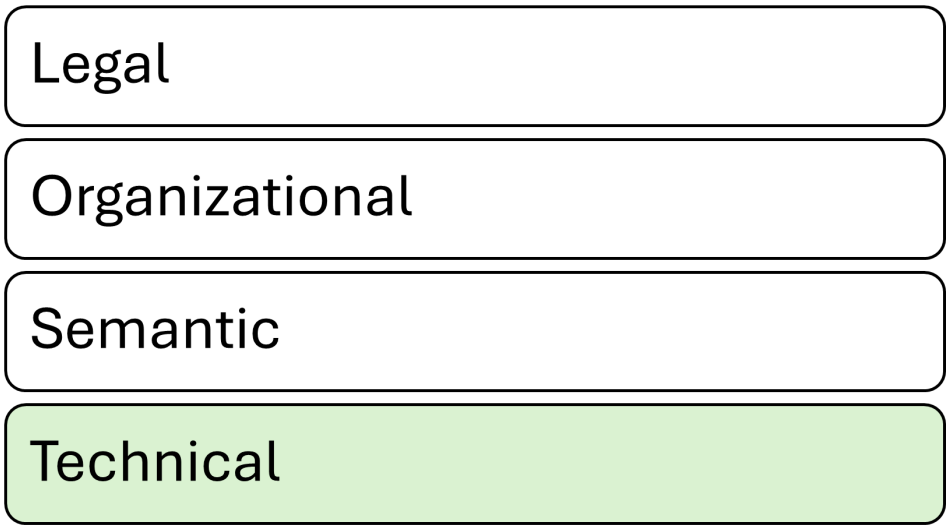

Figure 42: European Interoperability Framework

The DSGA will specify which version(s) of the DSP are mandatory for participants of a Data Space. This guarantees that at a technical layer participants will be able to interact in a Data Space.

In addition to the DSP, additional protocols, like identity and trust protocols can be defined to guarantee minimum technical interoperability within a Data Space.

If two Data Spaces mandate their participants to use the same protocols it will make the participation technically easier as participants might re-use the same technical components to access data in different Data Spaces, instead of having to maintain separate technical systems for each Data Space.

37.5.2 Semantic

An important part of interoperability is the semantic models used in a Data Space. This includes semantic models of the data that is to be shared within the Data Space as well as semantic models for the data that describes the Data Space itself (such as policies, participants, processes etc).

To successfully participate in a Data Space a participant needs to at least understand the semantic model used for policies within the Data Space. This is usually pre-defined through the DSGA and acceptance of the semantic model of the Data Space policies is very often going to be a pre-requisite to join the Data Space.

The semantic model of the Data Space's policies helps participants to understand the policies that can be negotiated in data sharing contracts. Without a common understanding of what individual policies mean and the expectations for their execution, it is not possible to participate in a Data Space.

The semantic model of the actual data being shared is not a mandatory required element, but it greatly enhances the value of the Data Space and the data shared within as it enables all participants to know what each data element means, and how it is constructed. For example, if a field refers to "country code", it is necessary that the participants in that particular data-sharing transaction know how that value is coded (e.g. ISO two-letter or three-letter abbreviation, some numeric representation, or whatever else has been chosen); if one participant is coding the United Kingdom as "UK" while another uses "GB", the danger of misalignment and miscommunication is obvious.

If two DSGAs negotiate and agree on the same semantic model for policies for their respective Data Spaces, it will greatly simplify the access of data in the two different Data Spaces.

37.5.3 Organizational

For a Data Space to be well governed a clear definition of organizational processes is required. Again, all participants in a Data Space will have to follow the same processes.

If two Data Spaces define the same organizational processes it will greatly simplify the participation in multiple Data Spaces.

37.5.3.1 Cross-Data Space interoperability If multiple Data Spaces define the same organizational processes, this will greatly simplify the participation in these Data Spaces.

37.5.4 Legal

As policies in Data Spaces might have legal consequences if they are not properly adhered to or executed it is important that participants understand the mapping of Data Space policies to legal constructs. This is already difficult enough to achieve for one Data Space, especially if participants reside – or operate - in multiple jurisdictions, it gets more complicated when a participant needs to access data from different Data Spaces. Just because a policy has the same semantic model in both Data Spaces doesn't mean that the policy has legal equivalency in both Data Spaces.

A participant in multiple Data Spaces will have the responsibility of keeping track of which data came from which Data Space and what the legal responsibility of handling this data is.

Agreements between legal organizations managing a Data Space can reduce the burden on the participants by agreeing on the legal equivalency of policies in both Data Spaces.

37.6 Interdependency models in Data Spaces

As shown above, the burden of ensuring interoperability and the adherence to all rules of individual Data Spaces is the responsibility of a participant. However, the complexity of achieving interoperability across multiple Data Spaces also greatly depends on how those are related.

The simplest model is probably a hierarchical direct-dependency of Data Spaces. In a larger Data Space, a smaller sub-Data Space could be created with additional rules, utilizing the governance model of the overarching Data Space, but introducing additional policies for the sub Data Space. E.g. think of an industrial Data Space where one participant wants to share data only with their direct suppliers instead of the entire Data Space. This can be realized as a separate Data Space or within the larger industrial Data Space by having an additional DSGA with additional membership policies and specifying additional semantic models and processes. In our example, one participant could specify that for a specific set of data other participants need to prove that they are suppliers of this participant and understand specific semantic models and processes provided by this participant. This can be regarded as a hierarchical relationship between two Data Spaces. In this case, interoperability should be straightforward to achieve.

Another model is Data Space peers. Two Data Spaces operating in different domains, but with a substantial overlap of participants, which also require data from both Data Spaces for many use cases. To reduce the burden on participants guaranteeing interoperability, the two Data Spaces might agree on the same requirements for protocols, semantic models, and also on organizational processes, including agreements on legal equivalency.

Last but not least, there is the case of completely unrelated Data Spaces where one or just a few participants have the need to access data from multiple Data Spaces. In this case the

burden of interoperability fully lies on the participants as they will need to be able to comply with processes in both Data Spaces and potentially will need to provide completely separate technical environments to access data in different Data Spaces.

No matter how Data Spaces are related and cross-data Space interoperability will be achieved, it is always going to be the responsibility of the participant to keep track of which data was acquired through which Data Space and what obligations came with it. Especially if use cases need to combine data from different Data Spaces sophisticated data management will be required.

37.7 Trust Frameworks and Trust Anchors

As the DSGA also defines which Trust Frameworks and which Trust Anchors can be used by participants within a Data Space with all the aforementioned interoperability facets also apply to Trust Frameworks and Trust Anchors. As it is very likely that Trust Frameworks and Trust Anchors will support multiple Data Spaces it is especially important that the applicable protocol versions and semantic models are clearly defined.

37.8 Improving Interoperability

As already established above, the main responsibility for interoperability in Data Spaces is with the participant, however, everyone involved in a Data Space can support interoperability by aligning with other parties.

Aspects of interoperability in Data Spaces can be achieved by utilizing common frameworks, models, standards, processes, or services, like Trust Frameworks. Those need to be mandated by participants of a Data Space as agreements in the Data Space Governance Framework executed and managed by the Data Space Governance Authority.

The agreements in the Data Space Governance Framework of one Data Space can and should be reused or acknowledged by other Data Spaces. This leads eventually to commonly adopted concepts and standards.

Recognizing the different levels of interoperability as described above, a general adoption or maturity model can be derived:

1. Agreements between 2 participants are highly tailored to certain use cases, but can provide a foundation for broader adoption.

2. Agreements in a group of participants are the foundation to increase interoperability by increasing the number of adopters.

3. Agreements within a common framework increase interoperability using the same technical or organizational framework.

4. Agreements between service providers support the broad adoption by providing interoperable solutions (as a service) to facilitate their reusability in the market and thus drive common interoperable functionalities.

5. General agreements in a Data Space as part of the Data Space Governance Framework implement default interoperability aspects which releases the participant from implementing alternative approaches or choosing between different approaches.

6 Agreements between different Data Space Governance Authorities establish bridges between data spaces.

- **Data Space to other Data Space**: negotiate legal equivalency of processes and rules between the two organizations.

- **Data Space to Trust Frameworks and Trust Anchors:** Align on mapping between policies and legal provisions and processes.
- **Data Spaces to DSGAs:** Align on governance models and organizational processes.
- **Trust Framework to other Trust Frameworks:** Share semantic models for policies and align on identity and trust protocols required.
- **Trust Framework to DSGAs:** Agree on standardized identity and trust protocols and a common set of semantic models.
- **DSGA to other DSGAs:** Share semantic models for policies and agree on functional processes.

38 Technical Agreements

39 Technical agreements

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 how to realize these concepts (see also the section on the goals of IDS in the IDS RAM).

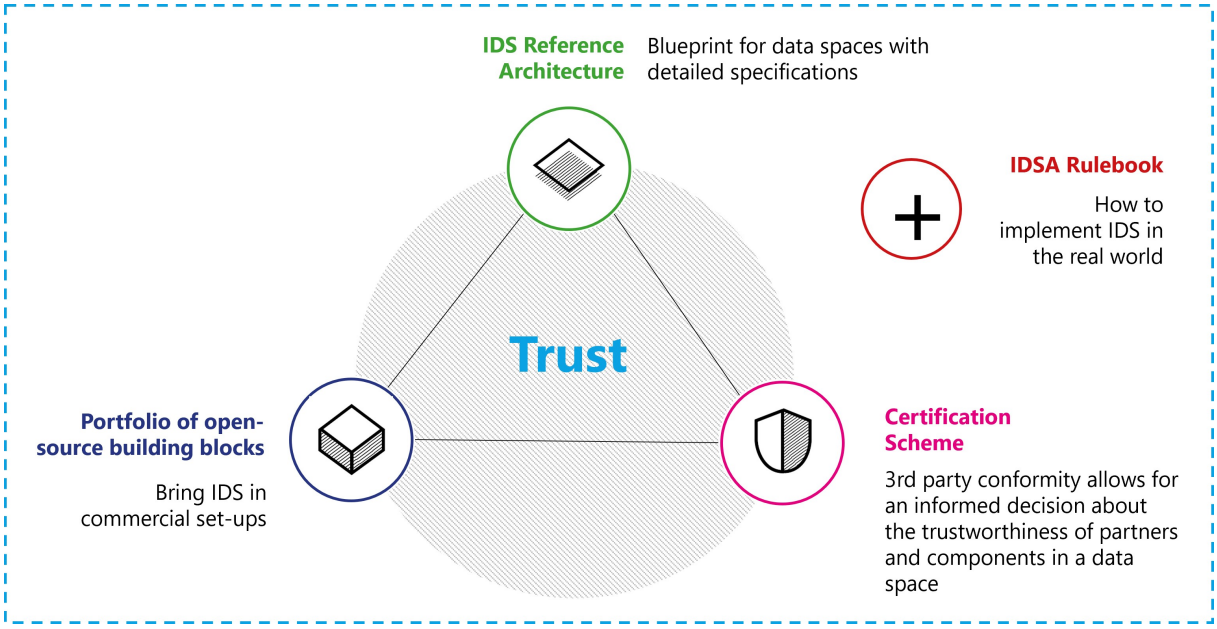

Figure 43: IDSA Magic Triangle

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 so-called magic triangle, which is extended with the portfolio of open-source building blocks, such as 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.

The mentioned IDSA assets have a defined release time to ensure consistency between them. In general, an IDS asset can be released after approval by the IDSA working groups and final approval by the technical steering committee. To achieve reliability for industrial use of the IDS assets, major releases that contain fundamental changes may be conducted once per year. For more details see 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-reference testbed	Fourth quarter of a year	Working group certification
IDSA Rulebook	Third quarter of a year	Working group Rulebook

40 IDS RAM

40.1 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 describes both the static and the dynamic aspects of the IDS constituents using data linkage principles. The system layer addresses 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 roles in the IDS.

The current version of the IDS-RAM that forms the basis for this Rulebook is V4.

41 IDS Specifications on IDS-G

41.1 IDS specifications on IDS-G

IDS-G 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 main 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

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.

42 IDS Certification

42.1 IDS Certification

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

- The certification scheme describes the operational model and roles in the IDS Certification.
- The rules of procedure include the formal outline of organizational processes
- Approval of evaluators
- Execution of evaluations
- The certification criteria 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.

43 IDS Testbed

43.1 IDS testbed (interoperability test)

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.

This includes:

- All evaluation facilities conduct transparent conformance tests in the „IDS reference testbed" based on the regulations from the certification working group and approved by the IDSA technical steering committee.
- All evaluation facilities assess compliance with the security requirements listed in the IDS criteria catalog based on tests derived from the criteria. Tests that can be conducted automatically are part of the test suite of the IDS-testbed.
- The evaluation facilities issue a certificate when conformance and security tests are 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 test suite provided.

44 Organizational Agreements

45 Organizational agreements

45.1 Certification

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

Functional requirements are an element of trust and are investigated from the functional perspective, clarifying 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 trust anchor and trust framework. The first term refers to the entity that issues certifications about an attribute, the second to the 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 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.

45.1.1 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 are based on a self-assessment mechanism, while the more advanced assurance levels require a third-party assessment of components or operational environments. This third-party compliance check is performed by the evaluation facilities, which are specifically approved to offer this service. The approval process is defined by the IDSA certification working group.

All the details on the IDS Certification scheme, the trust and assurance levels for component certification and operational certification, the certification criteria, and the process to approve the evaluation facilities are provided in Chapter 4.

46 Legal Dimension

47 Legal dimension

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

48 Regulatory Framework

48.1 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 result in a fragmented and incomplete regulatory framework. To address these shortcomings in data sharing and reuse, the EU Commission presented the "European strategy for data" in February 2020 describing the vision of a common European data space. The Commission has proposed different regulations (Digital Markets Act (DMA), Digital Services Act (DSA), AI Act on harmonised rules for data governance, data access and use as part of the EU's digital strategy.

Beside other regulations the Data Governance Act (DGA) entered into force on 23 June 2022 and will be applicable from September 2023 after a 15-months grace period. On 23 February 2022, the Commission proposed a regulation on harmonised rules for fair access and use of data, the Data Act Proposal (DA-E). With both acts the Commission aims to make more data available for use, by setting 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 reuse of publicly/held, protected data, by promoting data sharing through the regulation of novel data intermediaries and by encouraging the sharing of data for altruistic purposes. It aims to make public sector data more widely available for local businesses, researchers and communities for the development of innovative data-driven services. A specific focus is on the public sector data which is subject to legal restrictions and thus 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.

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. 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. A contractual framework is needed, providing clarity on rights and remedies regarding accessing, processing, sharing, and storing of data in order to limit misuse. 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.

Beside these specific Acts, further legal aspects to consider when sharing data, including antitrust/competition, data protection and security, copyright, patents/Intellectual property. The regulatory development may 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 to implement the upcoming legislative agenda of the EU. Providing guidance to future-proof specific problems entails an assessment and combination of various regulatory tools, contractual models, design principles, and organizational structures. To this end, the below ***four-pillar data governance framework*** outlines a "legal anatomy" of data governance consisting of the following:

1. the substantive rights and obligations related to data transactions (rights to data)
2. the contractual dimension
3. the organizational aspects
4. the technical implementation.

Beside the own responsibility of participants in a decentralized organization, IDSA discusses and aligns on legal matters with other initiatives. Coordination with other initiatives on the legal dimension is all the more important as often (and by its nature) most legislation needs 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 established a legal framework task force to discuss regulatory developments and legal topics as well as to organize the collaboration and contribution of IDSA members regarding the legal dimension.

49 Legal Agreements and SITRA Rulebook

49.1 6.2 Legal Agreements & SITRA Rulebook

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

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 their data sharing context. In a decentralized organization 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 needs (template approach).

Considering IDSA's focus on other dimensions and the importance of alliance with other initiatives, IDSA takes an "adopting & consolidation approach" to the contractual framework. Therefore, IDSA did not invent an own set of legal agreements for IDS participants to be used instead opting to suggest an already established contractual framework modified regarding IDS specifics.

SITRA's rulebook for a fair data economy provides tools for networks where organizations can share data and create 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 data network. Sitra published the first version of the rulebook for a fair data economy in 2020 and has updated it several times since. The rulebook model is backed by Sitra's longterm 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.

The basic principles of the Sitra rulebook align well with the goals of this IDSA Rulebook:

Sovereignty

IDS has made sovereignty of the data owner its most important 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 data terms of use, in which the data provider can decide to whom it grants access and under what conditions it releases the data for use by others in the data network. **Trust** Enabling trust is the first of the foundational IDS concepts. Trust is encompassed in the Sitra Rulebook primarily through the balance between the sovereign data providers and data users building new business with the data. For instance, 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 upon. By default, the data already distributed may still be used by the data user after termination, but the provision of new data is seized. The balance is also found in the clauses defining the boundary between data and derived material in a way that it fits the context and needs of the data provider and the data user. The Sitra rulebook also includes 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.

50 Contract templates for IDS

50.1 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 regularly updated reflecting new developments.

The IDS contract framework will not duplicate all components of the SITRA rulebook. The full set of SITRA rulebook templates are intended to be used in the creation and set-up of a data space, including governance models of the data ecosystem. These may not be necessary for the purposes of the 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 domain-specific dataset terms of use templates or more detailed components for cross-continent data sharing or privacy. If the IDS contract framework requires modifications to the SITRA rulebook's terms and conditions, they will be proposed also to Sitra's workgroup to maintain compatibility and to avoid different versions of terms and conditions.

51 Summary and Outlook

52 Summary and outlook

This IDSA Rulebook version 2 recognizes the growing need for structural approaches to accessing and sharing data while maintaining data sovereignty. The use of guiding principles helps identify solutions for this growing market. It includes the understanding of the current European regulation and legislation but does not stop there as data spaces are meant to be international.

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

The analysis of the legal framework for data spaces is ongoing and subject to continuous debate and will be part of this Rulebook in future versions.

IDSA and its partners are supported by the Data Spaces Support Centre (DSSC) which will offer additional guidance.

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

53 Annex 1: AI and Dataspaces: Shaping the Future of Trusted, Decentralized Intelligence

54 AI and Dataspaces: Shaping the Future of Trusted, Decentralized Intelligence

54.1 Introduction

The convergence of Artificial Intelligence (AI) and Dataspaces marks a pivotal moment in the evolution of data ecosystems that strongly benefits from recent developments in Trusted Data Sharing. As organizations seek to unlock the full potential of their data, the emergence of decentralized, context-rich data ecosystems—known as dataspaces—offers a new paradigm for secure, collaborative, and trusted data sharing and use. This paper explores the technological, business, and governance dimensions of AI in dataspaces, positioning them as the foundation for next-generation digital trust and innovation.

54.2 The Evolution of AI: From Learning to Reasoning

AI has entered its third wave, with machine learning and generative models (such as LLMs) now mainstream. The focus is shifting from data-driven learning to advanced reasoning, where AI agents must operate on real-time, context-rich data. The bottleneck is no longer just access to data, but the ability to reason with it in dynamic environments. Retrieval Augmented Generation (RAG) architectures combine LLMs with additional data sources to enhance reasoning. They allow AI agents to operate across distributed data sources, leveraging external tools and resources for complex tasks. Dataspaces provide both, access to data and the essential context, moving the field from "prompt engineering" to "context engineering". The technologies are a great match for each other as both have two important technical attributes in common: decentralization and asynchronous operations that enable autonomy and agency!

54.3 Dataspaces: The Architecture of Trust and Collaboration

Dataspaces are higher-order orchestration technologies that separate the data plane (data transfer) from business logic. They enable trusted, decentralized negotiation of access to data and APIs, supporting data sovereignty, interoperability, and reduction of business risk through attribute-based trust mechanisms. Participants—represented by software agents—negotiate sharing contracts, ensuring that data is shared with a set of agreed policies and permissions and rules guiding it's use. Key features include: - **Trust Contexts:** Dataspaces create environments where data sharing is governed by robust contracts and policies, reducing business risk. - **Decentralized Management:** No central data vault; participants retain autonomy and agency and can belong to multiple overlapping dataspaces. - **Legal and Regulatory Alignment:** Compliance with frameworks like the EU Data Act is integral.

54.4 AI Agents in Dataspaces

In general, there are two overarching scenarios of how AI Agents could leverage dataspace technologies. It's kind of a bit like the chicken and egg problem: which one was first? Is it from the perspective of a dataspace that has AI Agents acting on behalf of participants or is it an AI Agent acting on behalf of an organization that needs to join a dataspace or use dataspace protocols and processes to access and use data?

54.4.1 Dataspace First

AI agents operate within the boundaries of an existing dataspace and leverage pre-negotiated rights and contracts, streamlining data sharing and compliance. Participants in a dataspace have already agreed on data sharing contracts and AI Agents are being issued tokens to execute those contracts. One could distinguish between two patterns: - Dataspace Data Planes are used to transfer data from one participant to the other. Once all the relevant data is collected in one organization its AI agents can act on the data locally, as long as it respects the usage policies associated with the data. This is probably the easiest to implement but comes with several disadvantages, e.g. the potential staleness of data. - The use of MCP in Dataspace Data Planes. This would effectively wrap the access to data in a thin façade that enables the use of the MCP Protocol. The data provider would build a data plane that hosts a MCP Server and provides data access according to the data sharing contract. The data consumer would have an MCP client that is aware of data usage policies and guarantees that policies of the data sharing contract will be respected. This would allow more complex access scenarios than the first pattern.

54.4.2 Agent First

Agents discover data offerings that other organizations are publishing through dataspaces but don't yet have existing access. Depending on the rules of those dataspaces the Agents will need to first branch out into a human workflow to join the dataspace and negotiate the required data sharing contract. Alternatively, if the data provider supports direct negotiation the Agent can leverage dataspace protocols to automate the negotiation for data access without requiring its organization to manually join the dataspace. To negotiate access ad hoc, Agents can leverage dataspace protocols using decentralized claims and policy mapping for dynamic, cross-organizational attribute access control to be granted permission to shared data and to agree on data usage contracts.

54.5 Protocols and Governance

AI agents are becoming critical actors within data ecosystems, requiring new protocols for secure, trusted communication. Emerging standards such as the Model Context Protocol (MCP) and Agent-to-Agent (A2A) protocols are maturing, enabling agents to interact with databases, computational systems, and each other. Dataspaces bring a set of new protocols, Dataspace Protocol (DSP) and the Decentralized Claims Protocol (DCP) which enable a decentralized, attribute-based access control system for negotiation of data sharing contracts, governing access and usage permissions. Although those protocols are being developed separately, they fit together perfectly, solving for different problems on the path to providing AI Agents with high value data in a controlled but highly automatable and scalable way.

- **MCP**: Standardizes interactions between AI Agents and data resources, managing permissions and access on a technical machine to machine level
- **A2A**: Facilitates agent-to-agent communication, supporting federated and collaborative AI scenarios.
- **DSP**: Standardizes interactions between organizations. It enables the negotiation of data access and usage policies, as well as a high level control of data sharing activities.
- **DCP**: Moving beyond traditional identity providers, agents can prove compliance via verifiable claims, streamlining access and trust.

While MCP focuses on a token-based access control to individual resources DSP and DSP focus on the higher level trust creation that ultimately leads to the issuance of a token for resource access. When an AI Agent working on behalf of Organization A needs data from Organization B the two organizations can use Decentralized Identities (DIDs) and Verifiable Claims (VCs)

to exchange information about each other that leads to trustworthiness (without the need to establish a common identity provider). Trust between two organizations can be established when one organization defines policies for trustworthiness and the other organization provides evidence—through verifiable claims—that these policies have been met. By verifying these claims, both organizations can build confidence in the relationship, ensuring that trust is not assumed but demonstrated and substantiated.

DSP is used to negotiate a data sharing contract that contains details about the nature of the data, the type of data technology used, semantic models, provisioning requirements for resources and also usage controls which Organization agrees to adhere to. Once such a contract is negotiated and approved it needs to be executed. This is where MCP comes into play. Through DSP and DCP the AI Agent will be handed information on where to locate the MCP Server, which access token to use and what usage is permitted. It can then instantiate the MCP client, connect to the data or API to fulfill its task. The resource protected through a decentralized attribute-based access control can be a data set, but also an API or even another AI Agent. Same as the negotiation of data sharing contracts an AI Agent access contract could be negotiated. The data plane for that contract then would have to be an A2A protocol implementation.

54.6 Semantic Interoperability and Compliance

As it will be virtually impossible to create a single globally standardized semantic model for data sharing contracts, AI can help in the interpretation of different models and thus aid in the fulfillment of a data sharing contract. AI's ability to map different semantic models can in principle reduce the need for manual standardization, enabling automatic policy negotiation and compliance verification. Dataspaces often leverage domain-specific knowledge graphs and ontologies, providing the context layer essential for meaningful AI-based interpretation and integration. Shortcomings, however, may arise due to lack of robustness and reliability of AI in properly interpreting and mapping the involved semantic models, especially in automated dynamic, and possibly safety- and security-related, scenarios. This can lead to unforeseen violations of compliance requirements for AI systems, lowering their overall level of trustworthiness.

54.7 Strategic Recommendations

1. **Invest in Protocol Standardization:** Support the development and adoption of MCP, A2A, DSP and DCP to enable secure, scalable AI integration.
2. **Prioritize Governance and Trust:** Establish clear frameworks for identity claims, and compliance. Leveraging AI for automated policy evaluation.
3. **Enable Semantic Mapping:** Use AI to bridge semantic gaps, facilitating interoperability across diverse data sources and domains.
4. **Foster Consortia and Ad-Hoc Collaboration:** Design data offers and dataspace processes to support both structured and dynamic data sharing scenarios.

54.8 Conclusion

AI and dataspaces are converging to create decentralized, trusted ecosystems for intelligent data sharing and reasoning. As AI agents become more autonomous, robust protocols, governance models, and semantic interoperability will be essential. Organizations that embrace these innovations will lead to digital trust, compliance, and business agility.

55 Introduction

56 Introduction

The emergence of dataspaces as a key enabler for trustworthy data sharing has introduced a new class of data architecture principles. Within this landscape, the **International Data Spaces Association (IDSA)** provides a foundational Rulebook that defines the core principles, roles, and capabilities required to establish and operate such environments. However, turning these high-level principles into actionable, technical guidance requires a dedicated architectural framework. **This is the purpose of the Reference Architecture Model (RAM)**.

The RAM is the central compendium of technical documents that operationalizes the concepts defined in the IDSA Rulebook. It maps abstract governance models, data usage policies, and protocol specifications into concrete technical (logical) components, interfaces, and behavioral patterns. It does so with a clear focus: enabling interoperability, scalability, and conformance without prescribing a single implementation path. In other words, the RAM is not a blueprint but a design space—a structured but flexible guide that supports diverse requirements while preserving the integrity of the IDSA dataspace model.

This document is for system architects, software engineers, and infrastructure designers who are tasked with building or integrating components within a dataspace to enable business-driven data ecosystems. If you're looking to understand what makes a connector IDSA-compliant, how to build interoperable and integrable services, or how to maintain trust and policies in a decentralized environment—this is your technical guidance.

Central to the RAM are specifications such as the **Dataspace Protocol (DSP)** and the **Decentralized Claims Protocol (DCP)**. These protocols define how components communicate, how identities are exchanged and verified, and how policy-conformant data discovery and transfer is achieved. The RAM highlights these specifications, covering components, message formats, interaction sequences, and binding details. Further, it describes in depth how they are integrated with key capabilities like identity management, observability, data discovery, contract negotiation and secure data transfer. Details of such capabilities are maintained within dedicated documents under the purview of IDSA. The IDSA RAM document integrate and connect these documents to provide a comprehensive overview and a single source of truth for architecture models according to IDSA specifications.

To support real-world applicability, the RAM organizes its content into two major sections:

Capability Mapping: This section lists the essential capabilities enabled through dataspaces—such as data discovery, policy enforcement, usage control, identity resolution, and observability. Each capability is analyzed from a technical perspective, detailing how it can be implemented within a compliant dataspace. The descriptions are aligned with the IDSA Rulebook and maintains references to the foundational concepts to highlight the strong relation between the two documents. The RAM content is however grounded in system-level detail, interactions pattern, documents expected behavior, and provides guidance on integration patterns of infrastructure and other technologies.

Architectural Best Practices: Recognizing the diversity of business and technical requirements across domains, the RAM does not enforce a singular architecture. Instead, it presents validated patterns and warns against known anti-patterns, guiding implementers through the architectural decisions while setting up or operating a dataspace. Whether the goal is to create a lightweight edge connector, operate a multi-tenant marketplace, or integrate with existing enterprise systems, the RAM aims on outlining the architectural considerations and trade-offs—while maintaining compatibility with the IDSA model.

The RAM is intentionally neutral with respect to implementations. It refrains from endorsing any specific codebase or vendor product. Where useful, illustrative code snippets and configuration examples are included to clarify complex concepts and show practical realizations. For this, available open source projects (such as those maintained under the Eclipse Dataspace Working Group (EDWG) initiative) are referenced.

Importantly, the RAM is a *living* document due to continuous updates of IDSA technical documents this document refers to. Therefore, rather than locking into static version cycles, it evolves incrementally with latest releases of technical documents. Changes of these documents are managed through the corresponding GitHub repositories, with full traceability of modifications and clear visibility into the rationale behind decisions. Periodic release tags of the RAM document however will provide stable points of reference, allowing contributors and adopters to align their work with a consistent snapshot of the evolving model.

56.1 Contributions

Provide an overview on how to contribute to the document and how to get involved in the IDSA Architecture Working Group

56.2 Terminology

Provide a terminology in line with other IDSA (at best just link to one consistent)

57 Context

58 Relation to other IDSA Documents

The International Data Spaces Association (IDSA) provides a comprehensive body of documentation that serves as both a strategic compass and a practical guide for building interoperable, trusted dataspaces. IDSA documents follow a clear top-down structure: starting with vision and principles, moving to governance and requirements, narrowing into thematic clarity, and concluding with technical realization. This structure ensures coherence across conceptual, operational, and technical domains, enabling organizations to confidently adopt dataspace principles at scale.

Level	Document	Purpose
Vision	Manifesto	Defines core principles and motivation
Governance	Rulebook	Establish requirements and roles
Thematic guidance	Focus topics	Explore evolving topics in depth
Technical Architecture	RAM	Enables implementations

58.1 Manifesto – The IDSA North Star

At the top of the IDSA document hierarchy stands the IDSA Manifesto, a concise yet powerful declaration of purpose. It describes IDSA's ambition to shape a trusted global data economy founded on sovereignty, trust, and interoperability. The Manifesto serves three essential functions:

- Sets the shared aspiration: to enable data sharing ecosystems where organizations retain control over their data and use it responsibly for collective innovation.
- Establishes guiding values: trust, fairness, interoperability, and sovereignty, which frame all other IDSA documents.
- Calls for collective action: encouraging industry, research, and policymakers to collaborate toward Trusted Data Sharing based on open standards and decentralized architectures.

Unlike detailed guidance documents, the Manifesto is deliberately high level. It articulates "why" dataspaces matter and builds a shared identity and strategic direction for the IDSA community. All subsequent documents—including the Rulebook, Focus Topics, and the Reference Architecture Model (RAM)—are grounded in the principles laid out here.

58.2 Rulebook – From Principles into Practice

Guided by the ideals of the Manifesto, the IDSA Rulebook answers **"WHAT"** needs to be done to translate vision into concrete requirements and governance models. It supports the creation, operation, and growth of data spaces by distinguishing mandatory requirements from optional, value-adding practices. Its scope spans technical, commercial, and legal dimensions:

- Common technical guidance, including functional requirements and specifications.
- Recommendations for applying IDSA technical artefacts and for alignment with partner frameworks.
- Operational guidance for collaboration, roles, and processes that enable data space ecosystems.
- Perspectives on implementing and complying with international legal and regulatory obligations to facilitate trusted, cross-border data sharing.

The Rulebook acts as the normative foundation of IDSA. It does not specify implementation technologies but clearly defines conditions for trust—usage control, contractual assurance, and transparent operations—thus linking data value creation with accountability.

58.3 Focus Topics – Depth on Key Challenges

While the Manifesto inspires and the Rulebook governs, documents for individual focus topics dive deeper into specific adoption within dataspaces. These concise thematic modules ensure that the documentation can evolve with a fast-changing environment without overloading core documents. Current topics include:

- Identity & Trust – decentralized identity management and verifiable credentials.
- Interoperability – semantic, organizational, and technical interoperability models.
- Observability – monitoring dataspace operations while respecting sovereignty.
- Agentic AI and LLM – integration of MCP and dataspace in the context of agentic web

Each Focus Topic is anchored in the Rulebook and consistent with the Manifesto, providing reusable patterns and best practices. Their modular structure ensures efficient maintenance and allows new topics to be integrated as technology and regulation evolve.

58.4 Architecture Document – Technical Realization

While previous documents define what to implement, the Reference Architecture Models (RAMs) explain **"HOW"** to build it. It is the central technical compendium that transforms more abstract governance concepts into implementable architecture. The RAM introduces:

- Interaction models between logical components
- Protocol specifications like the Dataspace Protocol (DSP) and Decentralized Claims Protocol (DCP).
- Capabilities for identity, data discovery, policy enforcement, contract negotiation, and secure transfer.
- Integration patterns that support interoperability without imposing a single technology stack.

The RAM is not prescriptive software architecture; it is a design framework that accommodates diverse implementation paths. It supports system architects and developers by connecting high-level requirements from the Rulebook with concrete deployment scenarios. Its two core sections—Capability Mapping and Architectural Best Practices—make it an indispensable engineering guide for building sovereign, trusted data ecosystems.

59 Architectural principles

60 Architecture Principles

This chapter provides insight on the main technical concepts of dataspaces. It highlights the protocols DSP and DCP, their messages, state machines, sequences, and bindings

60.1 Cataloging

Provide insights on advertising data assets

60.2 Contract Negotiation

Provide insights on contract negotiation and agreements

60.3 Data Transfer

Provide insights on the actual data transfer within a dataspace

60.3.1 Control Plane

Explain the responsibilities of the control plane via data transfer

60.3.2 Data Plane

Explain the responsibilities of the data plane via data transfer

60.3.3 Policy Enforcement

Explain the policy enforcement capabilities and shared responsibilities of dataspace and data management services

60.4 Observability

Link to the observability document, interpret from an architectural pov

60.5 Credentials and Claims

Link to the identity document and the trust document, interpret from an architectural pov

61 Architectural Patterns

62 Architecture Pattern and Guidelines

This chapter provides more concrete architecture guidelines on how to design different architecture patterns within a dataspace. Beside the introduction and explanation, trade-offs are highlighted.

62.1 Data Space Governance Authority

Explain the DSGA from an architectural pov

62.1.1 Federated or Central

Provide insights on federated or central DSGA and corresponding trade-offs

62.1.2 Decentral

Provide insights on decentral DSGA and corresponding trade-offs

62.2 Catalogs

Explain Catalogs in the context of dataspaces from an architectural pov

62.2.1 Federated or Central (Marketplace)

Provide insights on federated or central catalogs and corresponding trade-offs

62.2.2 Decentral

Provide insights on decentral catalogs and corresponding trade-offs

62.3 Observer

Explain Observer from an architectural pov

62.3.1 Federated or Central Escrow

62.3.2 Decentral

Provide insights on decentral observability and corresponding trade-offs

63 Outlook

64 Outlook

Provide an outlook of currently discussed topics within the IDSA Architecture Working Group

65 Focus Papers

66 RAM 5 structure

- Introduction
- Context
- Architectural principles
- Architectural Patterns
- Outlook

67 Glossary

68 IDSA Glossary

This draft document is created to ensure consistent terminology across IDSA's documents. The definitions are aligned with the ISO DIS 20151 (to be published at the time of writing this document) and the Dataspace Protocol (2025-1 release) where possible. Additional notes are provided where different terms are used for the same concept across the two sources. Please note this is just an initial draft to provide a starting point. The structure and approach could be updated in the next steps.

68.1 A

68.1.1 Agreement

A concrete Policy associated with a specific Dataset that has been agreed between the Provider and Consumer. It is a result of a Contract Negotiation defining the Policy agreed to for a Dataset. Please also see data sharing contract

(Source: ISO/IEC DIS 20151)

68.2 C

68.2.1 Catalog

A collection of entries representing Offers that are advertised by a Provider.

(Source: Dataspace Protocol)

68.2.2 Catalog Protocol

A set of allowable Message Types that are used to request a Catalog from a Catalog Service.

68.2.3 Catalog Service

A Participant Agent that makes a Catalog available and accessible to Participants.

68.2.4 Connector (Data Service)

A Participant Agent that performs Contract Negotiation and Transfer Process operations with other Connectors, by implementing Dataspace Protocols. It produces Agreements and manages Dataset sharing.

68.2.5 Consumer

A Participant that requests access to an offered Dataset.

68.2.6 Contract Negotiation

A set of interactions between a Provider and Consumer that establish an Agreement. It is an instantiation of the state machine of a Contract Negotiation Protocol. An outcome of a Contract Negotiation MAY be the production of an Agreement.

68.2.7 Contract Negotiation Protocol

A set of allowable Message Type sequences defined as a state machine.

68.3 D

68.3.1 Dataset

Data or a technical service that can be shared by a Participant.

68.3.2 dataspace

data space

governance framework and supporting services to build trustworthiness and enable data sharing through an agreed set of Policies, semantic models, protocols and processes

[SOURCE: ISO/IEC DIS 20151]

68.3.3 dataspace governance authority role

DSGA role

set of activities provided by one or more parties that establishes, governs, manages and enforces the technical policies and business rules of a dataspace

[SOURCE: ISO/IEC 20151 to be published]

68.3.4 dataspace participant

participant

party that is acting in a dataspace participant role

Note 1 to entry: By being accepted to be a participant in the dataspace, the party agrees to the governance arrangements and therefore the policies of the dataspace. [SOURCE: ISO/IEC 20151 to be published]

Please also see Participant definition sourced from DSP.

68.3.5 dataspace participant role

participant role

set of activities within a dataspace for the purpose of data sharing or related activities Note 1 to entry: Related activities can include auditing or observing roles that do not include data sharing or governance activities. [SOURCE: ISO/IEC 20151 to be published]

68.3.6 data policy

human and machine-readable set of rights and obligations regarding access and use of data [SOURCE: ISO/IEC 20151 to be published]

68.3.7 Dataspace Protocol

A set of Messages and Message sequences that enables the interaction between Participant Agents in a Dataspace. This may require additional concepts, which are not in the scope of the specification itself.

68.3.8 Data Transfer Protocol

A set of rules and conventions that dictate how data is transmitted over a network by defining the format, error handling, and flow control. Examples include HTTP, FTP, MQTT, and AMQP.

68.3.9 data sharing

Access to the same data by more than one authorized entity Note 1 to entry: Use of the data can be synchronous or asynchronous. Note 2 to entry: Data can be shared, for example, (i) by allowing access to, or the execution of operations over, the original dataset, or (ii) by giving a copy of the data to the interested entity. Note 3 to entry: The way in which data is shared fundamentally influences the available controls and the statements needed in a data sharing agreement. [SOURCE: ISO/IEC 23751:2022[4], 3.7, modified - removed 'or processing of']

[SOURCE: ISO/IEC 20151 to be published]

68.3.10 data sharing contract

formal and legally binding agreement between dataspace participants containing policies, terms and conditions for data sharing Note 1 to entry: Data sharing contracts usually contain information about access to data, including its metadata, and data use. Note 2 to entry: A data sharing contract is usually much more specific than a data sharing agreement which is often broader and often at an organizational level. [SOURCE: ISO/IEC 20151 to be published]

Please also see Agreement for a related definition sourced from DSP.

68.3.11 data use

Handling or dealing with data for a specific purpose [SOURCE: ISO/IEC 5207:20245, 3.30, modified – Note 1 to entry removed] [SOURCE: ISO/IEC 20151 to be published]

68.4 G

68.4.1 governance

Human-based system comprising directing, overseeing and accountability [SOURCE: ISO/IEC 38500:2024[6], 3.3] [SOURCE: ISO/IEC 20151 to be published]

68.4.2 governance framework

Strategies, policies, decision-making structures and accountabilities through which the organization's governance arrangements operate [SOURCE: ISO/IEC TR 38502:2017[7], 3.1] [SOURCE: ISO/IEC 20151 to be published]

68.5 M

68.5.1 Message Type

A definition of the structure of a Message.

68.6 O

68.6.1 Offer

A concrete Policy associated with a specific Dataset.

68.7 P

68.7.1 Participant

A member of one or more Dataspaces that provides and/or consumes Datasets. It registers Participant Agents that perform tasks on its behalf.

Please also see dataspace participant definition sourced from to be published ISO 20151

68.7.2 Participant Agent

A technology system that performs operations and interactions in a Dataspace on behalf of a Participant, such as publishing a Catalog or engaging in a Transfer Process. It is a logical construct and does not necessarily correspond to a single runtime process. While most interactions take place between so-called Connectors, some interactions with other systems are required.

68.7.3 Policy

A set of rules, duties, and obligations that define the terms of use for a Dataset.

68.7.4 Profile

A restriction or subset of a specification that enforces every occurrence of an externally defined class to be conformant with the original definition.

68.7.5 Provider

A Participant that offers a Dataset.

68.8 T

68.8.1 Transfer Process

A set of interactions between a Provider and Consumer that give access to a Dataset under the terms of an Agreement. It is an instantiation of the state machine of a Transfer Process Protocol.

68.8.2 Transfer Process Protocol

A set of allowable Message Type sequences defined as a state machine.

68.8.3 trust

Decision by an entity to assume that a product, service or entity will behave as expected for a given circumstance

[SOURCE: ISO/IEC 20151 to be published]

68.8.4 trustworthiness

set of verifiable evidence that can be used to form trust

[SOURCE: ISO/IEC 20151 to be published]

69 Standards and specifications

70 Standards and external sources

70.1 Specifications

- Dataspace Protocol

70.2 Standards

- ISO/IEC DIS 20151

71 Downloads

72 Downloads

This page provides exports of the complete Knowledge Base in **PDF** and **DOCX** format.

72.1 Latest exports

- Knowledge Base PDF (latest)
- Knowledge Base DOCX (latest)

72.2 Versioned exports

Each push to `main` produces a versioned export as well:

- `knowledge-base-YYYYMMDD-RUN-SHA.pdf`
- `knowledge-base-YYYYMMDD-RUN-SHA.docx`

If you need a historic version, you can also find it in the GitHub Actions run artifacts.

73 About

74 About

This knowledge base integrates curated documentation into a single destination:

- Build system: **MkDocs Material**
- Deployment: **GitHub Pages (Actions deployment)**
- Quality gates: **Markdown lint**, **broken link check**, **strict MkDocs build**

Provenance. Content under **Knowledge** is assembled from:

- `International-Data-Spaces-Association/IDSA-Rulebook` (`documentation/`)
- `International-Data-Spaces-Association/RAM5` (`docs/`)
- `International-Data-Spaces-Association/glossary` (`Glossary/`)

All external content is **copied during CI only** and **never committed** back to this repository.