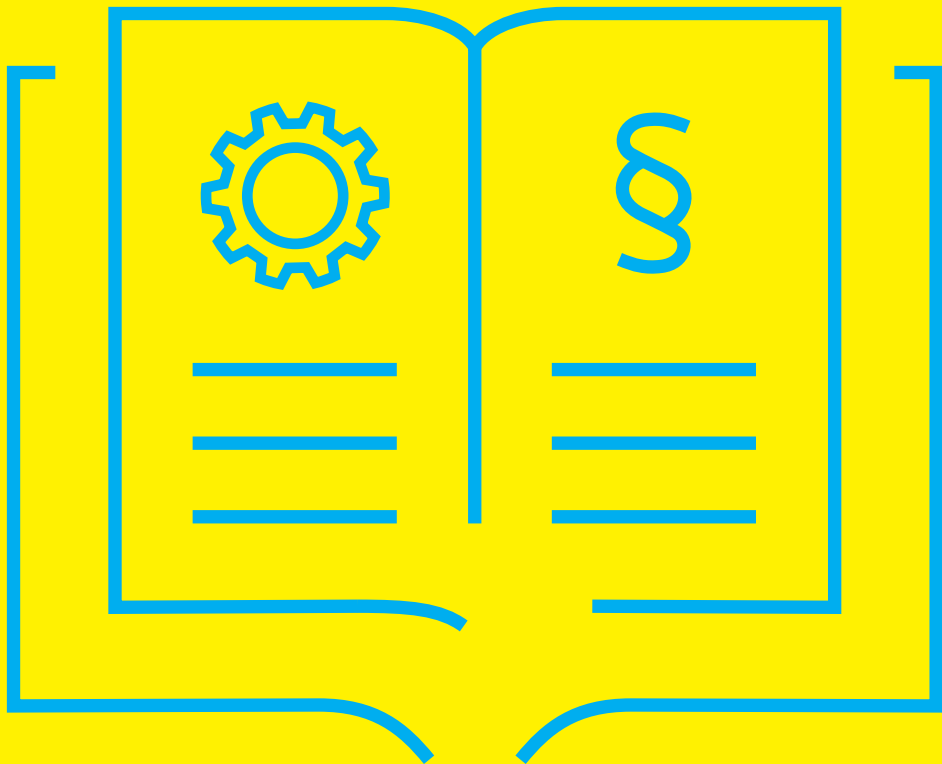


White Paper | Version 2026-2

IDSA Rulebook



- ☐ Position Paper of members of the IDS Association
- ☐ Position Paper of bodies of the IDS Association
- ☐ Position Paper of the IDS Association
- ☒ White Paper of the IDS Association

Publisher

International Data Spaces Association
Emil-Figge-Straße 80
44227 Dortmund
Germany

Copyright

International Data Spaces Association,
Dortmund 2026



Editor

Sebastian Steinbuss, IDSA

Authors & contributors

Sebastian Steinbuss, IDSA
Ilknur Chulani, IDSA
Marko Turpeinen, 1001 Lakes
Olaf-Gerd Gemein
Peter Koen, Microsoft
Markus Spiekermann, Huawei
Jim Marino, Metaform Systems
Anil Turkmayali, IDSA
Andreas Krimbacher, nexyo
Mario Holesch, IDSA
Petteri Kivimäki, Nordic Institute of
Interoperability Solutions

Cite as

Steinbuss S., IDSA Rulebook, International
Data Spaces Association, July 2026
<https://doi.org/10.5281/zenodo.21704653>

The IDSA Rulebook is updated on an ongoing basis. You can find the most current version here: <https://kb.internationaldataspaces.org/downloads/knowledge-base-latest.pdf>

Contents

1	Introduction	4
1.1	Who should read this Rulebook?	4
1.2	Goals and scope of the IDSA Rulebook	4
1.2.1	The purpose and scope of the IDSA Rulebook	4
1.2.2	Goals of the IDSA	5
1.3	Relationship with other organisations, projects & initiatives	6
1.3.1	How do initiatives relate in the field of Data Spaces?	6
1.3.2	International Data Spaces Association (IDSA)	6
1.3.3	ISO/IEC 20151: A Common Foundation	7
1.3.4	Dataspace Protocol (DSP) and Decentralized Claims Protocol (DCP)	7
1.3.5	Eclipse Dataspace Working Group (EDWG)	7
1.3.6	Eclipse Dataspace Components (EDC)	8
1.3.7	Technical Compatibility Kit (TCK)	8
1.4	The Data Space Landscape	8
1.4.1	Data Space Connector Report	8
1.4.2	Data Spaces Radar	9
1.5	How to contribute	9
2	Guiding principles	10
3	What is a Data Space?	11
3.1	Autonomy and Agency	12
3.2	Communities	12
3.2.1	International Cross-Border Data Sharing	13
3.3	Technology	14
3.4	Business Models	14
4	Understanding Participants, Roles and Layers in Data Spaces	15
4.1	Layers of a Data Space	15
4.1.1	Technical Layer	15
4.1.2	Economic Layer	15
4.1.3	Legislative Layer	15
4.2	Clarifying the Concept of Roles	15
4.3	Clarifying Policies and Contracts Across Layers	16
4.4	Distinguishing Data Spaces from Trusted Data Transactions	16
4.5	Participation and Representation	17
4.6	External Actors	17
4.7	Implications for the Rulebook	18
4.8	Conclusion	18
5	Roles	19
5.1	Technical Role - Participant	19
5.2	Data Space Governance Authority	19
5.3	Common Business Roles	19
5.3.1	Data consumer	19
5.3.2	Data provider	19
5.3.3	Service Provider (intermediary, value-adding services)	20
5.4	Summary	21
6	Dataspace Governance Authority	23

6.1	Definition	23
6.2	Core Principle - Decentralization	23
6.3	Implementation of the DSGA	23
6.4	Additional considerations for DSGA design	24
7	Decentralization Principles for Data Spaces	25
7.1	The benefit of decentralized data spaces	25
7.2	Maximising Participant Autonomy and Agency (Sovereignty)	25
7.3	Protocols for Interoperability: Leveraging DSP and DCP	25
7.4	Roles in the data space: Governance Authority and Participant	26
7.5	Trust Frameworks and Credential Management	26
7.6	Advanced Business Functions: Mapping to Participant Roles	27
7.7	Global data space mesh	28
7.8	Use Case Segmentation	28
7.9	Decentralization as the default	28
8	Trust in Data Spaces	29
8.1	Definition	29
8.2	Non-Assumptions	29
8.3	Separation of Concerns	30
8.4	Claims	30
8.5	Trust establishment mechanisms	30
8.6	Trust is a Runtime Property	31
8.7	Trust and Policy Interaction	31
8.8	Revocation and Trust Withdrawal	31
8.8.1	Revocation	31
8.9	Failure Modes	31
8.10	Interoperability Constraints	32
8.11	Governance Implications	32
8.12	Explicit Invariants	32
9	Dataspace Trust Frameworks	33
9.1	Definition	33
9.1.1	Core Principles	33
9.1.2	Trust Establishment and Maintenance	33
9.1.3	Governance Coupling	34
9.1.4	Implementation Considerations	34
10	Understanding the planes of a data sharing solution	35
10.1	Control Plane	35
10.2	Data Plane	35
10.2.1	Examples of Data Planes	35
10.3	Data Management Plane	36
10.3.1	Typical Functions of Data Management	36
10.3.2	Typical Functions of Data Governance	36
10.4	Application Plane	36
10.4.1	Key Characteristics	37
10.4.2	Exemplary Functions	37
11	Functional requirements for a Data Space	38
12	Achieving Participant Autonomy and Agency	39

13 Foundational concepts of a data space	40
14 Establishing trust	41
14.1 Establishing trust is the fundamental reason for data spaces to exist	41
14.2 Increasing trust lowers risk	41
14.3 Attribute-based trust	41
15 Attributes & Claims	43
15.1 Participant information	43
16 Policies	45
16.1 Policy Design	45
16.1.1 Membership Policies	45
16.1.2 Access policies for data discovery	46
16.1.3 Contract Policies	47
16.1.4 Usage Policies	47
16.1.5 Policies for segmentation	48
17 Data space membership	50
18 Identity	51
18.1 Attributes & self-description	51
19 Data space participation	52
20 Data sharing	53
20.1 Contract negotiation	53
20.1.1 Data sharing execution	54
21 Creating a data space	56
22 Data Spaces Interoperability - How to achieve Interoperability within a Data Space and across multiple Data Spaces	59
22.1 Motivation for interoperability	59
22.2 Guiding principles for Data Spaces	59
22.3 Interoperability Models	60
22.4 Interoperability Standards	61
22.5 Interoperability facets in Data Spaces	62
22.5.1 Technical	63
22.5.2 Semantic	63
22.5.3 Organisational	63
22.5.4 Legal	64
22.6 Interdependency models in Data Spaces	64
22.7 Trust Frameworks and Trust Anchors	65
22.8 Improving Interoperability	65
23 Catalog	67
23.1 Catalog(s)	67
23.1.1 Access policies	68
24 Observability	70
25 Vocabularies	72
25.1 Optional functions	73

26 Marketplaces	75
27 Processing services	76
27.1 Data escrow, data trustee	76
28 Connector	77
29 AI and Dataspaces: Shaping the Future of Trusted, Decentralized Intelligence	78
29.1 Introduction	78
29.2 The Evolution of AI: From Learning to Reasoning	78
29.3 Dataspaces: The Architecture of Trust and Collaboration	78
29.4 AI Agents in Dataspaces	78
29.4.1 Dataspace First	79
29.4.2 Agent First	79
29.5 Protocols and Governance	79
29.6 Semantic Interoperability and Compliance	80
29.7 Strategic Recommendations	80
29.8 Conclusion	80
30 Onboarding Pattern for Decentralized Data spaces	81
30.1 Joining a data space	81
30.2 Selecting a hosting model and provider	81
30.3 Onboarding to a data space	81
30.4 Obtaining Credentials from external data space Trust Frameworks	83
30.5 Using credentials in the data space	84
31 Summary and outlook	85

1 Introduction

1.1 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 organisations, 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 decentralized 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 and legally compliant data sharing; to manage business risk and contractual governance; and to implement technical architectures that preserve participant autonomy and agency.

1.2 Goals and scope of the IDSA Rulebook

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

Reading guide for normative language used in this Rulebook:

- “must” indicates a mandatory requirement for conforming implementations or for compliance where explicitly stated.
- “should” indicates a recommended best practice that is expected unless a justified exception is documented.
- “may” indicates an option or permissive practice.
- “recommended” and “preferred” are non-binding guidance intended to reduce risk or improve interoperability.

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 programmes like the Data Free Flow with Trust (DTFF)

1.2.2 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 standardisation bodies and open-source initiatives to harmonise 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 Data Space (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 honour 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 organisations 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.

1.3 Relationship with other organisations, projects & initiatives

1.3.1 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 organisational 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 standardised, 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 organisations, 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.

1.3.2 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, organisational, and technical considerations into a coherent vision. At the technical level, IDSA emphasizes the importance of a common core for specification and standardisation. 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 standardisation 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.

1.3.3 ISO/IEC 20151: A Common Foundation

One of the major cornerstones in formal standardisation 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.

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

The Dataspace Protocol (DSP) and Decentralized Claims Protocol (DCP) are two key specification projects that operationalise 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.

1.3.5 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 standardisation 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.

1.3.6 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.

1.3.7 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.

1.4 The Data Space Landscape

1.4.1 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.

1.4.2 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.

1.5 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 and also take our Code of Conduct into account.

2 Guiding principles

The IDSA Rulebook is based on a set of fundamental principles and underlying values, described in detail in the Manifesto of international data spaces. Key aspects are the autonomy and agency of participants in dataspace and their governance, as well as the responsibilities of participants in a data space.

Additionally several core principles apply to the guidance presented in the IDSA Rulebook:

- **No reinventing the wheel:** proven processes & technologies are used wherever possible.
- **Integrate existing systems:** integrating data spaces into existing systems is necessary to create end-to-end use cases and well functioning data ecosystems.
- **Use existing standards:** align data space guidance with international standards and specifications, re-use technical standards, and established processes wherever possible.
- **Industry and domain independent:** data spaces are applicable as a foundational concept and form a horizontal standard component for data ecosystems.
- **Easy to use:** reduce friction of implementing data spaces with a focus on portability and replicability.

IDSA applies four key governance principles:

- **Accountability:** Parties must be answerable for their actions and commitments, provide verifiable evidence of compliance, and maintain clear governance contact points for remediation and escalation.
- **Transparency:** Governance processes, policies, and operational procedures should be documented, discoverable, and auditable to enable informed participation and public scrutiny where appropriate.
- **Fairness:** Rules and operational processes must avoid undue advantage, ensure equitable treatment of participants, and provide impartial dispute resolution mechanisms.
- **Responsibility:** Participants and governance bodies are accountable for implementing policies, enforcing obligations, and providing timely remediation where violations occur.

As a result, IDSA offers free use of IDS specifications and related open resources for all, open governance processes in which everyone can participate, and transparent decision making - preferably by consensus.

3 What is a Data Space?

The ISO/IEC 20151 Standard “Information technology — Cloud computing and distributed platforms — Dataspace concepts and characteristics” defines Data Spaces as:

“environment enabling trusted data sharing between participating parties, based on an agreed governance framework, along with an agreed set of policies, semantic models, standardised protocols, processes, and facilitating services”

Let's try to unpack that and have a look at what all of that means:

- **environment** - An environment is more than just the infrastructure on which a participating party operates their software agent (typically a Dataspace Connector). It includes technical and non-technical elements that are coming together to enable the creation of trust which reduces risk when sharing data.
- **trusted data sharing** - Higher trust between two parties sharing data reduces business risk in the process of data. This can include many facets: knowing the other party, understanding how they manage data, getting guarantees on how the data was created or how it will be used. Trusted data sharing enables two parties to share data with confidence, understanding the risk of that sharing and managing it well.
- **participating parties** - a data space consists of a community of participants which are following a common set of rules and have provided a minimum set of assurance about themselves to increase trust in them. Typically two participants of a data space operate directly, peer-to-peer without intermediaries. However, there are scenarios where intermediaries can be helpful to facilitate specific business use cases.
- **agreed 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 on Data Space Governance Authorities (DSGAs) explains the functions, responsibilities and interactions of such a framework in more detail.
- **agreed set of policies** - Within a data space there are multiple layers of rules which ensure the creation of trust between two participants. Those rules are expressed as policies, specifically data space membership policies, access policies, contract policies and usage policies. More details on policies can be found in the functional requirements section on Policies.
- **semantic models** - Sharing common semantic models on (1) the data that is to be shared and (2) the data that describes the data space itself (e.g. policies, participants, processes) between the participating parties greatly enhances the interoperability within the data space. Exchanging semantic models of the data shared further improves its use and enables value generation. Semantic Models are explained as well in great detail in their respective section of the IDSA Rulebook.
- **standardised protocols** - The incredibly rapid growth and widespread adoption of the WWW would not have been possible without common protocols, such as TCP/IP and HTTP, which enable software from many different vendors, servers at many different providers and browsers on different phones and computers to seamlessly work together. Data spaces are built on a similar set of foundational protocols that enable technical communication between participants, namely the Dataspace Protocol (DSP) and the Decentralized Claims Protocol (DCP).
- **processes** - Different communities will have different processes for onboarding and managing their community members. Having those processes clearly defined, governed and well managed enables smooth collaboration between participating par-

ties.

- **facilitating services** - Many data spaces will cooperate with external service providers to support their operations and make data sharing easier and more secure. These services may include onboarding, auditing, marketplaces, and others. The IDSA Rulebook outlines key service categories and governance principles but does not provide an exhaustive list. As data spaces evolve, new business models, new governance models and services will naturally emerge.

3.1 Autonomy and Agency

The most important aspect of data spaces is the **autonomy** and **agency** of a participant, commonly referred to as “digital sovereignty”. Autonomy refers to the ability to decide whether to share specific data, with whom, and under which conditions. Agency refers to the ability to execute those data sharing decisions in practice. This is only fully possible if the participant controls all technical elements required to participate in a data space. Any external service that is mandatory (that is, required by the DSGA or by law/regulation) in the negotiation of a data sharing agreement or in the execution of such an agreement — for example, where a central or federated catalog is the only means to discover data or a marketplace is the only service that enables contract negotiation — reduces autonomy and agency.

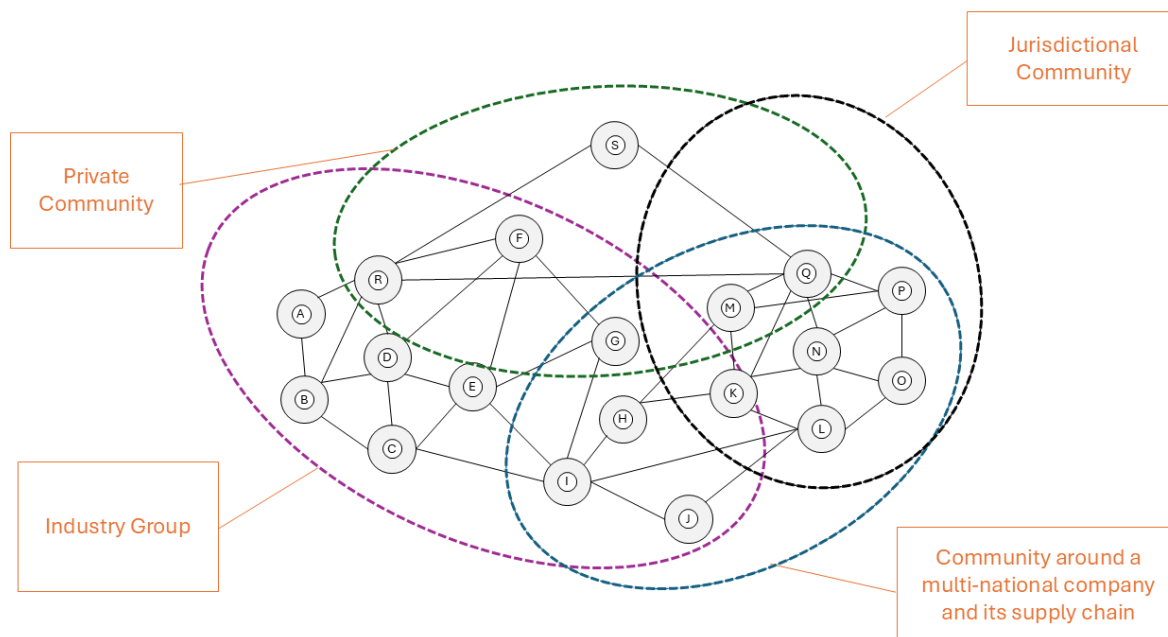
However, full autonomy and agency come at a price. The participating party needs to have the ability and technical means to control all technical, business and legal elements of participating in the data space and negotiating contracts. For many participants this will be too expensive or too cumbersome in relation to the value of the data shared. In those cases data intermediaries can be used which will perform data space access functions and perform the sharing or use of data. As each function that is provided by a service provider impairs the participants digital sovereignty it needs to be carefully weighted which functions can and should be handed over to a service provider.

Especially for smaller organisations participating in data spaces that primarily serve regulatory purposes, it may be preferable not to connect directly, but instead to use an aggregator or portal provider that shares data with the data space on their behalf.

The IDSA Rulebook will explain the functionality, mechanics and processes of a data space from the perspective of a fully autonomous participant with full agency. Where applicable additional explanation will be provided how intermediaries, aggregators and other service providers can be leveraged to trade digital sovereignty for ease of access and operation.

3.2 Communities

Data spaces are data ecosystems built on the concept of communities of trusted participants. There are many potential perspectives of what a community of participants can be. They can be based on a jurisdictional community (e.g. all participants are from the same country), industry group (e.g. all participants are energy providers, or health care providers), based on a common customers (e.g. a data space for all suppliers of a multi-national company), or any private community, for example founded to join forces on one very specific business use case.



Data spaces can overlap or be aligned hierarchically. As an example, the community of all European Healthcare Organisations is organised in the European Healthcare Data Space (EHDS), which can be segmented into smaller healthcare data spaces by country, while at the same time being segmented into separate data spaces by industry (e.g. pharma, hospitals, medical device manufacturers) and again be also segmented by a specific use case (e.g. cancer research, disease tracking).

No matter which community an organisation wants to participate in, it needs to have the technical means for participation. This includes, for example, operating a data space connector that supports the required foundational protocols and having knowledge of the relevant semantic models. In addition, the organisation needs to possess the appropriate credentials that prove its membership in the respective community.

Some communities will require significant effort to join (e.g. becoming part of a not-for-profit association, signing legal documents). Others may offer a more lightweight, low-friction onboarding process, for example when sharing open data for scientific research.

3.2.1 International Cross-Border Data Sharing

Special attention is necessary when the community consists of organisations operating under different jurisdictions as this might create legal conflicts in the policies required to share data. While some might be relatively easy to resolve (e.g. data sharing within the member states of the European Union), others will require more investments to resolve conflicts and clarify ambiguous rules (e.g. sharing data between a Chinese and a German company as part of a supply chain data space).

The IDSA Rulebook can't provide detailed guidance for every potential permutation of jurisdictional communities or industry groups. However, it will provide the tools necessary to understand how to think about the issues involved in designing the rule sets for different communities and how that impacts certain design and operational aspects of a data space.

Data spaces technology is agnostic to the custom rules of the communities. It always operates the same way by following the specifications for the two core protocols DSP and DCP.

Any customisation to the needs of specific communities starts with the semantic models, which are highly flexible and can accommodate any community need.

3.3 Technology

Every organisation participating in a data space is represented by a software agent which acts on behalf of the organisation in the data space. These software agent implementations are commonly referred to as Dataspace Connectors. The most basic Dataspace Connector offers API endpoints for data discovery, data sharing contract negotiation, data sharing orchestration and for the management of decentralized attribute-based claims about the organisation. This functionality is fully specified in the Dataspace Protocol (DSP) and the Decentralized Claims Protocol (DCP).

Based on those two protocols any variety of software agents can be built. Standalone, single server deployments, Connectors-as-a-Service at a cloud provider, software agents as a feature of a large enterprise software, AI Agents with Dataspace capabilities, and many more. The concept is open to future innovation and interoperability is guaranteed through the protocol specifications which embody the durable core concept of data spaces.

3.4 Business Models

Data spaces are agnostic to the business model(s) of the data ecosystem being enabled by them. It is possible to build data spaces that contain only free, open or altruistic data, others that are designed to share regulatory mandated data or others that are build around the idea of creating a marketplace for data. Any business model can be built on data spaces and with data space technology. Even platforms and aggregators can use data space technology to implement the need for a decentralized mechanism to negotiate resource sharing contracts.

Data spaces can also support more than one business model at a given time. e.g. same data that needs to be shared for free within the supply chain due to legal regulation can also be offered in a data marketplace to other companies at a price.

Data spaces are designed to foster future innovation in business models, data use, value creation and monetization.

The Internet and the World Wide Web have enabled new business models and monetisation strategies, creating significant value for the global economy and for individual users. Similarly, data spaces enable trusted data sharing, unlocking new insights, business models, opportunities, and value for all participants.

4 Understanding Participants, Roles and Layers in Data Spaces

Data spaces are multi-layered data ecosystems that rely on the seamless integration of technical protocols & components, 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 in different layers and across domains. This chapter establishes a clear separation between technical and non-technical interpretations of such concepts to support the development of interoperable data spaces.

4.1 Layers of a Data Space

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

4.1.1 Technical Layer

Encompasses the architecture and protocols (e.g., Dataspace Protocol (DSP), Decentralized Claims Protocol (DCP)) that facilitate trusted and interoperable data sharing. Software agents, representing the participants of the data space, form a decentralized mesh of autonomous nodes.

4.1.2 Economic Layer

Manages the services, interactions, and workflows that enable value generation. Alternative terms for this layer are also *Business* or *Operational Layer*.

4.1.3 Legislative Layer

Enforces rights, obligations, and regulatory compliance across participants. Alternative term for this layer is *Regulatory Layer*.

While at the technical layer all participants are equal the layers building on top can lead to segmentation into individual communities as described in the chapter What is a Data Space - Communities

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

4.2 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 Dataspace Protocol.

Note: In this Rulebook, “**participant**” denotes the technical role representing a legal organisation that operates software agents on behalf of that organisation. When the document discusses business or legal roles (e.g., provider, consumer, DSGA functions), those usages are explicitly qualified and belong to higher layers (economic or regulatory). This includes roles such as **data intermediary**, **marketplace operator**, **auditor**, or **service provider**, but also explicitly the **data provider** and **data consumer** role in the context of the business transaction.



Figure 1: Layers of Data Spaces

These business roles do not exist at the technical layer but are mapped onto the core participant role based on the actions performed and services offered.

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

4.3 Clarifying Policies and Contracts Across Layers

The terms **"policy"** and **"contract"** are also layer-dependent and should be interpreted explicitly in context:

- At the **technical layer**, policies and contracts are represented as machine-readable constraints and negotiated agreement artefacts used by software agents.
- At the **economic and legislative layers**, policies and contracts express legal and business intent, obligations, liabilities, and governance commitments between organisations.

The Rulebook therefore distinguishes between technical representations and legal/business meaning to avoid ambiguity and to support interoperability across implementations.

4.4 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 environments that enable participants to share data while ensuring their autonomy and agency. See the chapter What is a Data Space for details.
- **Trusted Data Transactions**, as under current standardisation in the European Commission's Standardisation 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 prioritise regulatory alignment and controlled environments.

While TDTs 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. Data spaces designed, built and operated according to the IDSA Rulebook enable the implementation of TDTs but don't mandate it. Also the implementation of a TDT is not dependent on the use of a data space. Both are independent concepts that can be used at the same time.

4.5 Participation and Representation

Participants in a data space are defined by their ability to automate the negotiation and execution of data sharing agreements via technical protocols. This has several implications:

- **Organisations**, not individuals, are considered technical participants. These organisations are represented by **software agents** capable of executing standardised data space protocols.

Note: For technical participation, "organisation" includes incorporated entities, not-for-profit bodies, public sector organisations, and registered sole traders or self-employed individuals acting as legal entities. Individuals may participate only when acting as a legal entity and if they are able to operate all required software agents directly or via an organisation that represents them (for example, a service provider).

- **Natural persons** shall interact with data spaces only indirectly through applications or services operated by organisations, and their participation shall be limited to the Economic and Legislative Layer. The services may include e.g. personal data intermediation (including consent management) or data wallet provision. Software agents may act on behalf of a natural person, but shall not impersonate or assume the identity of that person in the technical layer.
- As data spaces are fully decentralized and participants are responsible for their own **decentralized identity (DID)**, there are **no identity providers** in a data space. Participants provide proof of their identity to others through the use of claims (e.g. expressed through verifiable credentials) and not through a common identity provider. Access to resources is managed through authentication tokens issued directly by the participants sharing those resources.

4.6 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 implementing the standardised data space protocols. For example:

- A web service that hosts a data sharing ontology is not a participant but serves as an **external reference**.
- A dataspace trust framework provider usually acts as an external reference. However, it may act as a **participant** if it delivers services subject to data space governance policies provided through an implementation of standardised data space protocols.
- **Trust anchors, regulators** and similar legal entities may influence data transactions but do not participate directly unless they act through technical interfaces governed

by data space rules. Most commonly they provide external services which are referenced in the data space (e.g. common data lookup services, registries).

This model guarantees the adherence to the governance model and thus supports trust while preserving digital sovereignty of each participant and the integrity of technical interactions. Flexibility is created through processes in the economic layer which supports extensive customisation. The DSGA must not mandate central or federated services that restrict participant autonomy or agency without explicit justification; any such choice must be documented and accompanied by mitigation measures to preserve participant control.

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

4.7 Implications for the Rulebook

The Rulebook reflects these principles clearly:

- The only **technical role** is the **participant**, which may act as data provider, data consumer, or both.
- **Business roles** are supplementary and must be defined within the economic or regulation layers.
- **Clarity:** Visual representations and descriptions in the IDSA Rulebook 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.

4.8 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.

5 Roles

5.1 Technical Role - Participant

As established in the chapter on Layers the only technical role is the **Participant**. In this Rulebook, a **Participant** denotes the technical role that represents a legal organisation (not a natural person) and is capable of operating the required software agents; where an individual acts, they do so as a legal entity or via an organisation that represents them. However, there are many different roles possible in the economic and legislative layers of a data space. This chapter explains the most important ones.

5.2 Data Space Governance Authority

The Data Space Governance Authority (DSGA) is a special, functional role within a data space. It can be understood as the “legislative” of the data space by having the functional responsibility to define the rules and processes of the data space.

It serves a logical purpose within the Data Space Governance Framework but has no single definition of how it is implemented. To fully explain the role of the DSGA and not to distract from the Participant roles it is described in detail in its own chapter: Data Space Governance Authority(DSGA)

5.3 Common Business Roles

Business 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 processes. 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: data provider, data consumer, and intermediary services.

5.3.1 Data consumer

Data consumers are the recipients of a data sharing activity. In the IDSA Rulebook the data consumer is the party acting as the consumer of a data sharing contract.

The closely related term “data user” describes a natural or legal person who has lawful access to certain personal or non-personal data, and has the right to use that data for commercial or non-commercial purposes.

The data user and data consumer can be separate organisations but often will be two perspectives on the roles of the same participant, depending on the context.

5.3.2 Data provider

Data providers are the organisations that make data technically available for a data sharing activity. In the IDSA Rulebook the data provider is the party acting as the provider of a data sharing contract.

The closely related term “data holder” describes a natural or legal 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 data in accordance with applicable law.

The data holder and data provider can be separate organisations (e.g. in cases where intermediaries are being introduced) but often will be two perspectives on the roles of the same participant, depending on context.

Role mapping guidance:

- **Data Holder:** the legal owner or rights holder of the data (responsible for legal compliance and consent decisions).
- **Data Provider:** the entity that makes the data technically available (responsible for data access, availability, and technical controls).
- **When roles differ:** Contracts must explicitly allocate responsibilities (consent, liability, enforcement, and notification) and the DSGA may require that member self-descriptions include role assignments to make responsibilities discoverable.

5.3.3 Service Provider (intermediary, value-adding services)

In a data space optional service providers can offer optional capabilities to enable data sharing or to provide business services. 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. Their implementation must be representable as a participant at the technical level.

Intermediaries are services that are acting on behalf of participants in the data space. This can be a range of activities within the data space: negotiating a data sharing contract on behalf of a participant, implementing technical infrastructure on behalf of the participant and making decisions on behalf of a participant, thus impacting their digital sovereignty.

Such intermediaries may be regulated by local governments like the EU Data Governance Act in the European Union, which defines a specific “Data Intermediary” role. A detailed analysis can be found in the paper Reflections on the DGA and Data Intermediaries.

Intermediary usage guidance:

- **Optional:** Intermediaries may be used voluntarily by participants to simplify operations or reduce cost. When optional, their use is a matter of participant choice and does not reduce the baseline autonomy provided by the data space.
- **Mandatory (Regulatory):** If law or regulation requires an intermediary role, the DSGA must explicitly document this requirement and provide safeguards (e.g., audit rights, transparent processes, alternative providers where feasible) to limit negative impacts on participant autonomy.
- **Impact mitigation:** Where intermediaries are required or offer substantial control over interactions, the DSGA should require explicit documentation of the scope of delegated authority, revocation mechanisms, and participant redress processes.

Value-added service providers act as participants in the data space and therefore conform to the data space governance framework. Such services support value creation with a broad set of discovery, processing, and governance capabilities. Common examples and their expected functions include:

- **Data Discovery Services (catalogs, search engines, registries):** Provide standardised metadata, search interfaces, and registry functions that allow participants to locate datasets and service offerings. These services should provide interoperable metadata schemas, and support attribute-based-access-control visibility rules.

- **Lookup Services (vocabularies, shared information):** Maintain authoritative semantic resources such as controlled vocabularies, ontologies, and code lists that ensure consistent interpretation of attributes and data elements across participants. Lookup services should be versioned, discoverable, and accompanied by governance information describing scope and maintenance procedures.
- **Observability Services (audit, notary):** Offer verifiable logging, attestation, and auditing capabilities that help detect misuse, support dispute resolution, and provide forensic evidence. Observability services must balance provenance and transparency with privacy and minimisation principles.
- **Commercial Services (marketplaces, auctions, match-making):** Facilitate commercial discovery, negotiation, and transaction orchestration for data products and value-added services while operating within the governance and policy constraints of the data space. These services must not introduce mandatory central controls unless such centralization is explicitly justified and documented by the DSGA.

5.3.3.1 Potential roles from legal definitions Roles might also be described by legal regulation. An example of such regulation that defines roles in data sharing are the European Union 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)

5.4 Summary

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 affiliation to the layers.

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

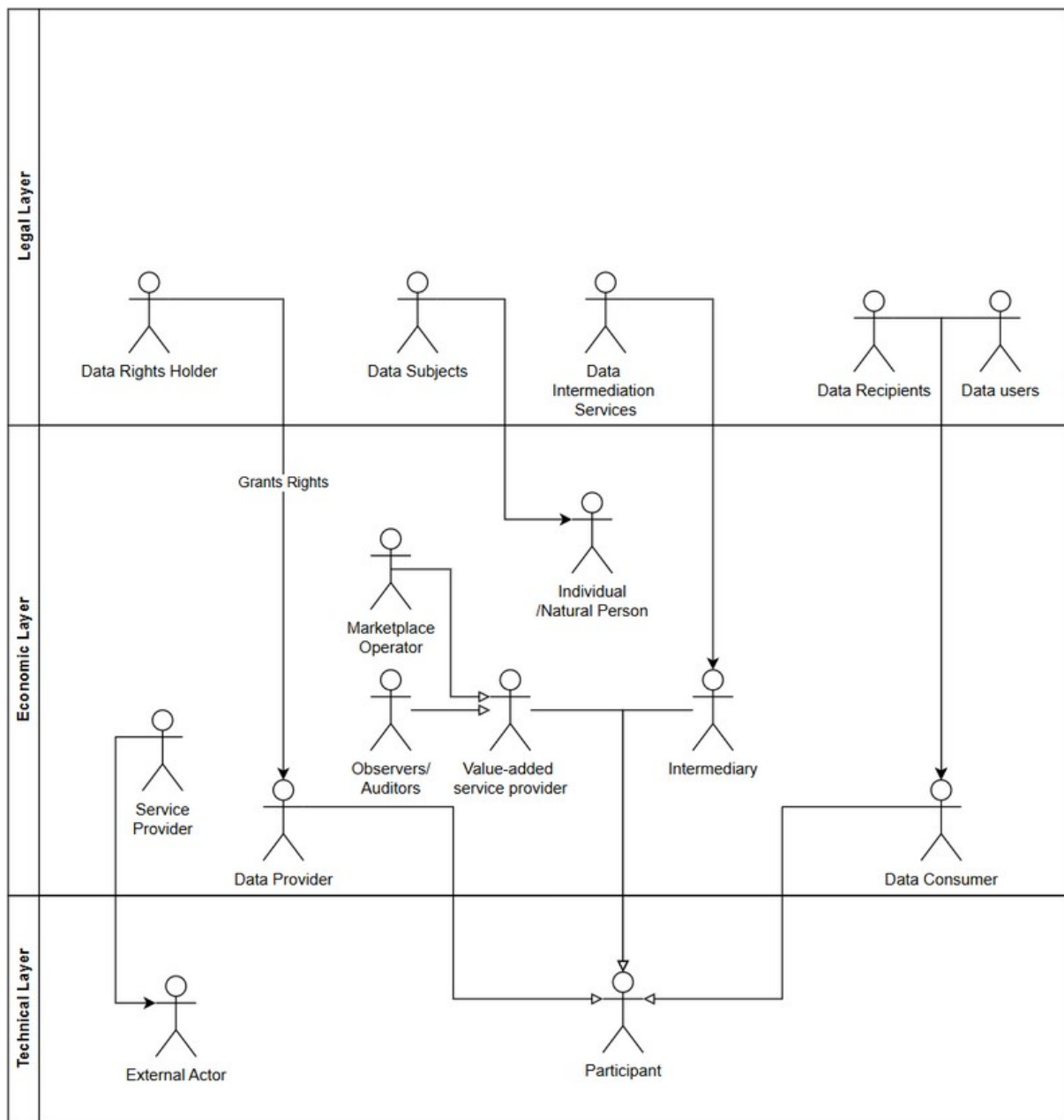


Figure 2: Overview on roles and their affiliation to layers in data spaces

6 Dataspace Governance Authority

6.1 Definition

A Dataspace Governance Authority (DSGA) is a functional role representing the collective governance function of a data space. It is responsible for providing the basic governance framework, defining which Dataspace Trust Frameworks (DTFs) are used for the data space, and where needed defining explicit business processes, policies, and semantic models to be used by participants in the dataspace.

The DSGA ensures interoperability through minimal shared semantics.

The DSGA is not by itself a runtime-enforcement entity. Normatively, it is expressed by the set of DTFs used within the dataspace together with any additional provisions, definitions, and processes specific to the individual data space.

While the DSGA is a logical role defining a governance model, this governance model needs to be enforced. There are multiple options how this can be achieved.

6.2 Core Principle - Decentralization

A DSGA should not prescribe mandatory central services or federations without providing clear justification and documented mitigation measures. The DSGA applies to all participants and defers responsibility for implementing the governance model and its enforcement authority across participants, using consensus-based reconciliation to resolve conflicts.

Important: An Operator of a value-added data space service is NOT a DSGA. They can operate a service that supports the implementation of the governance model provided by the DSGA.

A DSGA is a formal specification of the governance model and rules of a data space; it describes policies, trust frameworks, and operational processes. While the DSGA itself is not a deployed runtime service, it must be instantiated (operationalised) through one or more implementation patterns — for example participant-enforced controls, service-provider implementations, or an operations company — and the chosen instantiation must be documented, transparent, and explain how enforcement is achieved.

6.3 Implementation of the DSGA

In this context, an external **Oracle** is a participant-operated or independent service that provides verifiable, authoritative statements needed by participants to evaluate policies or claims (for example status, compliance, or reference data).

- **By the Participants:** each participant has full knowledge of the DSGA and can enforce the rules of the DSGA in interactions with other participants. Value-Added Services or external Oracles might provide additional information needed to implement the governance model defined by the DSGA.
- **Service Providers:** multiple service providers operate value added data space services or external Oracles that enable the enforcement of the governance model provided by the DSGA. The participants of the data space are responsible to choose which services providers to use and are free to reject interactions with participants where a common set of accepted service providers cannot be identified. An example of such a service can be the issuance of a data space participant credential.

- **Operations Company:** in a centrally governed data space the DSGA functions can be implemented by a single operations company which provides mandatory governance enforcement services. This creates centralized control and single points of failure and therefore impacts participant autonomy and agency.

The list above is illustrative, not exhaustive. Which implementation model is appropriate depends on the legal, business, and operational context of the specific data space.

6.4 Additional considerations for DSGA design

- **Minimal Semantics:** Interoperability is achieved through shared core concepts (e.g., participant identity attributes, data space operations), not comprehensive ontologies (e.g. semantic models of the data being shared in the data space) to reduce complexity.
- **Socio-Technical Coupling:** Governance rules need to integrate technical feasibility with legal participant agreements. Ambiguities in policy interpretation are resolved through defined rules in the DSGA or its referenced DTFs. DSGA can provide procedures and escalation paths, assuming reasonable consensus thresholds to resolve conflicts in the data space.

7 Decentralization Principles for Data Spaces

7.1 The benefit of decentralized data spaces

In an era defined by rapid digital transformation and complex data ecosystems, the architecture of data spaces must empower participants with maximum autonomy and agency to support their digital sovereignty. IDSA recommends decentralized data space architectures as the recommended approach because they best preserve participant control, equal treatment, and robust interoperability. Alternative architectures may be used where operational or regulatory constraints justify them, but any departure from decentralization must be documented and the impact on participant autonomy mitigated by the DSGA.

The core principles and practical advantages of decentralization in data spaces are well aligned with the ISO 20151 - “data space concepts and characteristics” standard.

7.2 Maximising Participant Autonomy and Agency (Sovereignty)

At the heart of decentralized data space architecture is the principle of participant autonomy and agency. Each organisation or entity operates independently, and is fully capable of deciding when to share which data assets with whom, under what circumstances and which rules apply to the usage of the shared data. Participants are managing their own credentials describing their identity, without reliance on an external party controlling their identity through a singular identity provider. They manage their interactions without centralized authorities or services. Where external or centralized components are required (for regulatory, operational, or other justified reasons), the DSGA must document the rationale and specify mitigation measures to preserve participant autonomy and choice. This approach protects digital sovereignty, allowing participants to decide how, when, and with whom their data is shared — critical for compliance, privacy, and strategic control.

7.3 Protocols for Interoperability: Leveraging DSP and DCP

Interoperability is essential for data spaces to function as collaborative, decentralized meshes of participants. With software agent (e.g. connectors) implementations leveraging Data Space Protocol (DSP) and Decentralized Claims Protocol (DCP) to facilitate seamless, standardised interactions between participants.

Technical interoperability between individual participants can be guaranteed irrespective of the data space which is operating as a governance and business context on top of the mesh of participants. These protocols enable software agents (e.g. connectors) to exchange data and services without vendor lock-in, ensuring that integrations remain flexible, secure, and future-proof.

Let’s revisit the mental model of layers of a data space. At the technical layer a data space consists of a decentralized mesh of individual nodes, which are acting with full autonomy and agency. Only in the higher layers of business and legislation processes the segmentation into separated trust contexts are manifesting. Such trust contexts can be “strong borders”, representing boundaries between data spaces in different nations, but also can be “weaker borders”, representing a segmentation into individual use cases within a data space. No matter where the segmentation takes place, all use cases are unified by the common, interoperable technology, founded on a solid base of the DSP and DCP.



Figure 3: Layers of a data space

7.4 Roles in the data space: Governance Authority and Participant

As described in the chapter on Roles Decentralized data space architectures define only two essential roles: the **Data space Governance Authority (DSGA)** and the **Participant**. The DSGA establishes rules and specifies which **Dataspace Trust Frameworks (DTFs)** will be used, who operates the accepted Onboarding Services and which mandatory business processes exist, while participants actively engage in the data space to negotiate data sharing contracts and execute existing agreements. Notably, service provider organisations can host governance and onboarding services, providing the necessary legal and business frameworks for onboarding and compliance.

Any business role within the data space, e.g. Provider, Consumer, Auditor, Marketplace, and many others can be built as a specialisation of the technical role of the participant.

There is no need for custom technical architectures or specialised protocols to satisfy those business roles and their requirements. No additional architectural components (e.g. central or federated catalogs, identity providers, etc...) are needed to create an operational data space. On the contrary, adding such special architectural components reintroduces centralization and fragility to the data spaces and becomes a single point of control and potential failure, very often resulting in performance bottlenecks or preferred attack points during cybersecurity events.

7.5 Trust Frameworks and Credential Management

Trust within the data space is governed by at least one **Dataspace Trust Framework (DTF)**. DTFs contain the rules that are fundamental to trust creation within the data space. An empty DTF/no DTF also qualifies as a DTF as no rules can be interpreted as data being shared with anyone without conditions (e.g.: Open Data).

DTFs can be built hierarchically by partial DTFs, external DTFs, DTF building blocks, etc. It is the responsibility of the **DSGA** and/or the participant to resolve any potential conflicts between the applied DTFs and to ensure a final, unambiguous set of rules. If a data asset is offered under two different rule sets, it shall be treated as two separate contract offers.

Instead of maintaining membership lists, the architecture relies on onboarding credentials — requested/issued and managed by participants themselves, checked and validated by onboarding services and signed by signatory services, the credential issuance service of the DTFs. This model ensures that each participant is responsible for their own credential lifecycle, promoting autonomy and reducing administrative overhead.

The model also assumes that the membership criteria set by the DSGA can be expressed as verifiable credentials, and that the dynamic nature of both membership criteria and credentials is taken into account.

Credential verification is handled on-demand, reinforcing the decentralized nature of the data space and minimising the risk of single points of failure or control.

7.6 Advanced Business Functions: Mapping to Participant Roles

Advanced business cases such as participant matching, observer roles, and data marketplaces can be easily mapped to the participant role. For example:

- **Marketplace/Matching Participants:** Multiple participants can provide a service that allows the discovery of and connection with other participants using catalogs and vocabularies, without central mediation. Each participant that wants to participate in a marketplace or a matching service shares their metadata through a data sharing contract with the Marketplace/Matching Service Provider Participant, who then in return will offer a data sharing contract with the potential matches. All within the rules of the data space, ensuring that autonomy and agency is preserved. Having multiple, independent such Service Providers will greatly enhance the freedom of choice and resiliency of the data ecosystem.
- **Observers:** Entities wishing to observe or audit interactions within the data space can do so by joining as participants with observer-specific credentials. Participants that are negotiating a contract that requires auditing can then both negotiate a data sharing agreement with the Observer participant to share their individual log files of the transaction. The Observer will provide the service of auditing and reconciling those log files and in return issue a data sharing agreement with the two participants where the results of the audit will be shared. Again, all perfectly within the rules and processes of the data space, fully preserving participant autonomy and agency.
- **Data Escrow Service Providers:** Data escrow operations are managed by participants acting as service providers, who are offering a trusted data escrow environment - a confidential compute environment where two or more participants can share their data and have computation being performed on the data without any of those participants ever having access to all data at once. The Data Escrow Service Provider participant guarantees the operation of the environment, and the distribution of results. With special encryption methods it can also be guaranteed that the Data Escrow Service Provider never gets to see the actual data. This is enabling joint analysis of data while ensuring the highest level of data privacy. E.g., in medical research scenarios or financial services.

When designing the data space business functions it is important to pay attention that the introduction of mandatory value-added services might introduce unwanted centralization or federation thus leading to undesirable concentration of control and accidental single point of failures/attack that can negatively affect the participants in the data space. It is highly recommended to enable an open market of competing value-added services to ensure higher resiliency and avoid centralization of control.

7.7 Global data space mesh

As decentralized architectures proliferate, a global mesh of data space agents (e.g. connectors) is a long-term target state and is beginning to emerge incrementally. Each data space maintains its own legal and trust boundaries, ensuring that governance and compliance are localised and context-specific. At the same time, the underlying technologies - agents, protocols, credential management systems - are reusable across multiple data spaces, maximising efficiency and reducing duplication. This mesh enables organisations to participate in multiple data spaces seamlessly, leveraging consistent standards and interoperable technologies.

7.8 Use Case Segmentation

Segmentation within the data space is achieved via use case specific credentials which might be issued by credential issuers that are providing specialised DTFs for a specific use case. For instance, in the automotive supply chain, manufacturers, suppliers, and logistics providers each hold credentials tailored to their role and use case. Use cases could be anything from specific business processes, regulatory requirements to smaller communities created by the supply chain of a specific company. This segmentation ensures that data access, sharing, and collaboration are precisely controlled, supporting advanced business models such as just-in-time delivery, quality assurance, and regulatory compliance.

7.9 Decentralization as the default

A fully decentralized data space architecture delivers unmatched benefits in participant autonomy, digital sovereignty, and interoperability. By aligning with ISO 20151, leveraging DSP and DCP protocols, and streamlining roles and credential management, organisations can build robust, flexible, and future-ready data ecosystems. As the global mesh expands and technology is reused across domains, the potential for innovation and collaboration grows exponentially. Segmentation via use case specific credentials ensures that each participant operates within precise trust boundaries and as effectively as possible, paving the way for the next generation of digital business.

8 Trust in Data Spaces

8.1 Definition

Trust in a decentralized data space is a situational, time-bound, and purpose-specific assessment of whether another participant can be relied upon to act within declared constraints for a specific interaction.

Trust is always created locally by each participant through its own evaluation of evidence, claims, and policies.

Trust is:

- **Not global:** Trust determinations are specific to a single data sharing contract and apply only to the parties that negotiate and execute that contract; they do not automatically extend to other contracts or to different counter-parties.
- **Not transitive:** A trust relationship established between two participants (for example, A trusts B) does not imply or create trust between either party and a third party (for example, B and C); each relationship requires its own independent assessment. Trust anchors can simplify evidence verification chains but do not create automatic trust transfer between participants.
- **Not reciprocal:** A unilateral decision by participant A to trust participant B does not compel B to trust A; mutual trust requires independent and explicit evaluations by each party.
- **Not permanent:** Trust assessments are time- and context-bound; changes in evidence, claims, policies, compliance status, or operational context may invalidate prior assessments and should trigger re-evaluation according to documented rules.
- **Not equivalent to identity, certification, or prior membership:** Identity credentials, certifications, and membership status can serve as evidence for trust but are not alone sufficient to establish trust without additional contextual evidence, policy alignment, and runtime verification.

8.2 Non-Assumptions

A conforming decentralized data space must not assume the following as sufficient or universal substitutes for trust. Each item below explains why it is insufficient on its own:

- **A central authority, broker, registry, or trust anchor:** Centralized services may provide useful evidence or convenience, but reliance on them creates a single point of control and does not replace per-contract, evidence-based trust decisions.
- **Pre-established federation membership:** Membership in a federation (infrastructure, data space, business, etc...) may indicate a baseline of trustworthiness but does not guarantee that a specific participant satisfies the policy or technical requirements of any given contract or context.
- **Static certification or onboarding events as sufficient for trust:** One-off certifications or onboarding checks are useful provenance, but they do not capture runtime behaviour, compliance drift, or post-issuance changes and therefore cannot alone assure ongoing trust.
- **Homogeneous policy languages or enforcement stacks:** Shared policy languages may aid interoperability, yet different parties may implement or interpret policies dif-

ferently; trust frameworks must tolerate heterogeneity and reconcile policy semantics.

- **That trust can be inferred solely from cryptographic identity:** Cryptographic identity proves control of keys, not adherence to policies or intent; trust requires additional attestations, claims, and evidentiary evaluation.

Any architecture that depends exclusively on the assumptions above is not representative of a fully decentralized data space and is likely to fail to provide robust, context-aware trust.

8.3 Separation of Concerns

Trust assessment must be confined to the control plane.

The transfer process defined by the Dataspace Protocol (DSP) involves two logical constructs: a control plane and a data plane. Their characteristics are explained in detail in the Planes section.

The data plane is agnostic to trust logic and trust state. Trust decisions may influence if, when, and how data plane interactions are initiated or continued.

8.4 Claims

Trust is derived by claim reconciliation at interaction time. Valid claim inputs include:

- Verifiable claims (representing evidence about identity, attributes, roles, attestations).
- Declared policies and obligations.
- Observed runtime behaviour from prior interactions.
- Third-party assertions with verifiable provenance (expressed as verifiable claims).
- Contextual factors (purpose, jurisdiction, risk class, data category).

By default, and unless a DSGA/DTF explicitly specifies an alternative, the absence of an input SHOULD be treated as negative (or unknown) for the purposes of trust evaluation; DSGAs and DTFs must document the chosen default and any exceptions.

8.5 Trust establishment mechanisms

Trust establishment in data spaces relies on the exchange and verification of claims, dynamic policy negotiation, and evidence collection.

Trust formation is an explicit, local process.

Each participant:

- Evaluates incoming claims and policies against its own constraints.
- Determines acceptability for a specific action and scope.
- Maintains its own trust state without global publication.

Trust formation must be repeatable and enable auditable by the evaluating participant. The audit logs might have to be shared through data space observability

Trust outcomes must be explainable in terms of accepted and rejected claims.

Trust mechanisms shall not require mandatory reliance on a single centralized or federated identity provider. The DSGA may define and accept multiple trust anchors (credential

issuers) as part of the Dataspace Trust Framework. Acceptance of a given anchor is a governance decision and does not grant structural control over the data space to that issuer. Policy and claims exchange uses interoperable, machine-readable protocols and formats. Centralized-only identity architectures are not aligned with the autonomy and agency objectives of decentralized data spaces.

8.6 Trust is a Runtime Property

Trust exists only while its assumptions hold. Therefore:

Trust can be continuously re-evaluated for long-running interactions. Material changes in claims, policies, context, or behaviour should trigger re-evaluation. Cached trust decisions should have explicit validity bounds. A participant MAY reuse a cached trust assessment for similar future interactions only if the scope, risk class, and validity period are explicit, conservatively limited, and documented; cached assessments must be re-evaluated on any material change to claims, policies, or context.

There is no concept of “once trusted, always trusted”. Participants should define verification frequency according to the sensitivity of the data and the risk of the interaction (for high-sensitivity assets require on-demand or per-execution revalidation; for lower-sensitivity assets periodic revalidation may be acceptable). Revocation signals must be handled promptly by participants; the DSGA or chosen DTFs should document expected propagation and verification intervals for different risk classes.

8.7 Trust and Policy Interaction

Policies do not enforce trust; they express expectations. Trust emerges from:

- Comparing declared policies with verifiable claims.
- Assessing whether obligations can realistically be met.
- Evaluating enforcement mechanisms offered by the counterparty.

Policy compliance must be treated as probabilistic unless directly observable. Trust decisions should explicitly capture residual risk.

8.8 Revocation and Trust Withdrawal

Trust can be revoked at any time. Triggers for withdrawal include:

- Claim revocation or expiry.
- Policy violations or non-compliance signals.
- Context changes invalidating prior assumptions.
- Inability to re-validate critical assertions.

8.8.1 Revocation

- Propagates to ongoing interactions.
- Can require termination, degradation, or isolation of data access.
- Doesn't depend on centralized revocation services.

8.9 Failure Modes

Architectures implementing Trust should explicitly handle:

- False-positive trust due to over-reliance on credentials.
- False-negative trust due to incomplete information.
- Asymmetric trust where only one side evaluates rigorously.
- Stale trust decisions in long-lived processes.
- Strategic misrepresentation of policies or capabilities.

Ignoring these failure modes renders trust claims non-credible.

8.10 Interoperability Constraints

To remain interoperable:

- Trust mechanisms must rely on minimal shared and discoverable semantics.
- Claim formats must be extensible and schematically loose.
- Trust logic must not assume shared policy languages, implementations or engines.

Interoperability emerges at the protocol boundary, not the trust model.

8.11 Governance Implications

Trust is inseparable from governance. Therefore:

- Dataspace Trust Frameworks (DTFs) must define acceptable trust evidence categories.
- Dispute resolution must acknowledge divergent trust assessments.
- Cross-domain interactions must tolerate incompatible trust conclusions.

Consensus on outcomes is optional; consistency of process is mandatory.

8.12 Explicit Invariants

A decentralized dataspace conforming to this model therefore must uphold:

- Local autonomy of trust decisions.
- Explicit trust scope and duration.
- Continuous re-evaluation and revocation.
- No mandatory central trust services.
- No implicit trust via participation alone.

These invariants take precedence over compatibility with prior data space models.

9 Dataspace Trust Frameworks

9.1 Definition

A **Dataspace Trust Framework (DTF)** consists of a set of policies and reconciliation mechanisms for claims, as well as business process definitions that enable data space participants to establish trust and maintain mutual assurance in data sharing interactions. Trust is treated as a dynamic runtime property, derived from verifiable claims and policy alignments, rather than static certifications or centralized attestations.

Key terms:

- **Claim:** A machine-readable, cryptographically verifiable assertion issued by a trusted authority about a participant or an asset (for example, an identity attribute, certification, or dataset provenance). Claims serve as primary evidence in trust evaluations and must include provenance and validity metadata.
- **Policy:** A formal expression of rules and constraints that govern data access, usage, and sharing. Policies must be expressed in machine-readable languages and define required attributes, permitted actions, obligations, and enforcement expectations.
- **Reconciliation:** The deterministic or negotiated process that aligns policies and claims between parties to determine whether a proposed interaction complies with both parties' constraints. Reconciliation may involve policy transformation, attribute mapping, or escalation to manual review.
- **Autonomy & Agency (Sovereignty):** The capability of a participant to control its data, make independent policy choices, select preferred service providers, and unilaterally suspend or revoke access to shared assets according to its governance and legal constraints.
- **Failure mode:** A class of scenarios that prevent trust from being established or maintained—examples include irreconcilable policy conflicts, compromised or unverifiable claims, or material deviations from declared behaviour—which should be handled by well-defined escalation and termination procedures.

9.1.1 Core Principles

DTFs operate on the assumption that dataspace are fully decentralized socio-technical systems, where technical protocols are coupled with governance mechanisms. Control plane activities (e.g., policy negotiation, claim verification) are separated from data plane operations (e.g., actual data transfer or access to a protected resource) to minimise coupling and enhance scalability.

Interoperability is achieved through minimal shared semantics, such as a common vocabulary for claims, rather than heavy global schemas. This allows for evolutionary changes without breaking existing implementations and supports the implementation of domain specific data spaces and DTFs.

9.1.2 Trust Establishment and Maintenance

Trust is established through iterative claim exchange and policy reconciliation as described in the Trust section. Summarisable by those important points:

1. **Claim Issuance and Verification:** Participants issue claims about themselves or their data. Verification occurs via cryptographic proofs (e.g., digital signatures) or

decentralized oracles (services that provide verifiable external assertions, such as status or reference facts), without reliance on central authorities.

2. **Policy Alignment:** Policies are expressed as logical constraints. Reconciliation uses a negotiation protocol to find mutually acceptable terms.
3. **Runtime Monitoring:** Trust is maintained through continuous monitoring of invariants, such as data usage logs or revocation signals. Violations trigger automatic revocation or escalation to human oversight.

Failure modes may include:

- **Policy Incompatibility:** If reconciliation fails, interactions are aborted with clear error codes.
- **Claim Compromise:** Detected via integrity checks; compromised claims invalidate trust chains.

When multiple Dataspace Trust Frameworks (DTFs) apply to an interaction, the DSGA must define explicit reconciliation rules. Absent explicit guidance, a conservative default is to require the intersection of constraints (the most restrictive applicable constraints) to be satisfied. The DSGA should also define escalation paths (e.g., arbitration, human review, or a voting mechanism) for irreconcilable conflicts and document the expected outcomes and timelines in the governance material.

9.1.3 Governance Coupling

DTFs integrate technical and governance layers:

- **Socio-Technical Invariants:** Rules must be enforceable either through technological means or business processes at both protocol and organisational levels (e.g., a policy requiring audit logs must have corresponding legal agreements and data sharing agreements for the audit log data).
- **Evolution Handling:** DTFs support versioned policies and claims.
- **Revocation Mechanisms:** Participants can revoke trust unilaterally.

Trade-offs:

- Decentralization increases resilience but complicates reconciliation.
- Minimal semantics reduce overhead but require a robust negotiation protocol.
- Dynamic trust enables adaptability but demands continuous verification resources.

9.1.4 Implementation Considerations

DTFs should be designed with common base standards:

- Use standards like DID (Decentralized Identifiers), VCs (Verifiable Credentials).
- Avoid assumptions of global information, or synchronous communication.
- Not assume the availability of centralized services or components (e.g., member registries)

This DTF definition invalidates legacy models relying on static certifications or central brokers, as they do not scale to decentralized data spaces.

10 Understanding the planes of a data sharing solution

A data sharing solution comprises of several layers:

- **Control Plane:** enabling core capabilities in a data space, for managing the discovery of available assets and the negotiation of data sharing contracts and orchestration of data sharing
- **Data Plane:** where peer-to-peer data sharing happens outside the data space, providing access to or transmitting data according to the policies agreed upon in the data sharing contract
- **Data Management Plane:** Organisational functions responsible for managing the lifecycle, quality, and governance of data within a data sharing solution; a capability out of scope of the dataspace, but very essential for the data space participants
- **Application Plane:** The user-facing layer that consumes and processes data from the Data Management Plane to deliver value to the participant

10.1 Control Plane

The control plane in a data space is the layer responsible for managing the discovery of available assets and the negotiation and orchestration data sharing interactions contracts.

It operates independently of the actual data access, focusing on establishing and maintaining the data sharing contracts that provide the requirements for the data sharing execution. The core capability of data spaces, the dynamic trust negotiation through policy reconciliation are executed in the control plane.

The control plane influences (orchestrates) but does not directly handle data flows (data plane), data lifecycle management (data management plane), or application logic (application plane).

10.2 Data Plane

The Data Plane is the actual technical data access technology. Its role is to provide access to or transmit data according to the policies agreed upon in the data sharing contract.

10.2.1 Examples of Data Planes

- **RESTful APIs:** A common data plane implementation using HTTP-based protocols for synchronous data retrieval.
- **Message Queues (e.g., AMQP or MQTT):** Asynchronous data transmission via publish-subscribe models.
- **Peer-to-Peer Protocols (e.g., IPFS or BitTorrent-like systems):** Decentralized file sharing without central servers.
- **Streaming Protocols (e.g., Kafka or WebSockets):** Continuous data streams with subscriptions. It handles high-volume, low-latency flows while maintaining separation from control plane governance.

These examples assume minimal shared semantics for interoperability. Implementations must account for failure modes, such as network failures or policy violations.

Sovereign capabilities depend on the specific implementation of a data plane and the underlying data transfer technology.

10.3 Data Management Plane

Once data has been shared it usually needs to be managed to enforce governance models and ensure the adherence to policies negotiated in the control plane.

The Data Management Plane encompasses the organisational functions responsible for managing the lifecycle, quality, and governance of data within a data sharing solution. This includes ensuring data integrity, compliance with contractual obligations, and usability across internal systems.

The Data Management Plane encompasses the functions responsible for the lifecycle, quality, and governance of data within a data sharing solution.

While it operates separately from data transmission (data plane) and orchestration (control plane), the Data Management Plane supports participation in the data space by integrating governance requirements into the organisation's internal processes and systems.

10.3.1 Typical Functions of Data Management

- **Data Ingestion and Storage:** Collecting, validating, and storing data from various sources, ensuring format consistency and metadata attachment. It enforces invariants like data provenance and traceability to support sovereignty.
- **Data Processing and Transformation:** Applying algorithms for cleansing, aggregation, or anonymization, while adhering to policies reconciled in the control plane.
- **Data Cataloging and Discovery:** Maintaining metadata registries for asset discovery, ensuring that policies from data sharing contracts are being enforced.

10.3.2 Typical Functions of Data Governance

- **Policy Enforcement and Compliance:** Monitoring adherence to governance rules, such as retention policies or usage constraints, through runtime checks.
- **Quality Assurance and Auditing:** Assessing data accuracy, completeness, and lineage, with audit trails for trust verification.
- **Risk Management and Data Sovereignty Protection:** Identifying and mitigating risks, including sovereignty violations, by enforcing participant controls. It prioritises conservative approaches, such as deny-by-default for sensitive data.

These functions are clearly separated from other planes. However, interaction and integration with other planes is necessary for building full governance and management capabilities.

10.4 Application Plane

The Application Plane is the user-facing layer that consumes and processes data from the Data Management Plane to deliver value to the user. It operates independently of data transmission (data plane), orchestration (control plane), and lifecycle management (data management plane), focusing on application logic and user interactions. It integrates with organisational governance mechanisms to ensure compliant use of data.

10.4.1 Key Characteristics

- **Data Consumption:** Retrieves processed data from the Data Management Plane, applying policies for access and usage in coordination with the data management plane based on the permissions and access controls associated with the identified user.
- **Service Provision:** Provides services such as analytics, visualization, and reporting to end-users, enabling decision-making and insights. These services are built on top of managed data, prioritising usability and performance while adhering to governance rules managed by the data management plane.
- **User Interaction:** Supports interfaces for human or machine users (agents), including dashboards, APIs, or automated workflows.

10.4.2 Exemplary Functions

- **Analytics:** Performs computational analysis on data from the Data Management Plane, such as statistical modelling or machine learning, to generate insights.
- **Visualization:** Renders data into graphical representations (e.g., charts, maps) for user consumption.
- **Reporting:** Generates structured reports or alerts based on processed data, integrating audit trails for traceability.

The Application Plane must maintain clear boundaries with other planes, interacting via standardised interfaces. It can support dynamic trust by propagating policy constraints, mitigating risks such as sovereignty breaches and manage user consent mechanisms.

11 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 decentralized architectures and deployment patterns to show how various solutions are enabled by the building blocks of data spaces.

Data sharing in a data space is not limited to sending data from one participant to another but can be more complex, like code to data scenarios. Fundamentally, all sharing of data consists of peer-to-peer interactions. All business scenarios consisting of multiple actors are built on peer-to-peer data contracts of two participants.

A data space adds value beyond individual data sharing by enabling collective trust and thus enabling complex, cooperative data services and applications. These capabilities have functional requirements which need to be included in the design of a data space.

Different business, regulatory, legal, or technical requirements necessitate different business processes, governance frameworks and solution approaches. Some data spaces might require specialised, value-added services to provide critical business services, or fulfil a regulatory mandate while others might be designed so their participants have a maximum level of autonomy and maintain agency over how to share their data.

The architecture and functional requirements of data spaces as depicted in the IDSA Rulebook supports a wide variety of design decisions while emphasizing a common technical foundation that guarantees interoperability and reduces friction for organisations that want to participate in many different data spaces.

As described in the chapters on Layers and Roles involved business roles are implemented through the technical participant in a data space.

This underlines the need for all parties involved in a data space and in the sharing, and usage of data to adhere to a common set of rules, the policies provided by the DSGA, and the data provider.

The concrete mandatory and optional functional areas are detailed in Foundational concepts of a data space.

12 Achieving Participant Autonomy and Agency

Participant Autonomy and Agency, commonly referred to as **Digital Sovereignty** starts with control over each participants own identity. Only participants that are in control of their own identity can be truly autonomous and have the agency to act in a data space.

The Manifesto of international data spaces articulates the fundamental principles. The functional requirements section of the IDSA Rulebook details the mechanisms required to achieve this goal

Any central or federated control impairs the participant autonomy & agency, reducing their ability to act within the data space.

IDSA recommends decentralized architectures and decentralized identities (DIDs) as the preferred approach to maximise participant autonomy and agency. In practice, the DSGA may define acceptable forms of identity evidence, trust anchors, and onboarding processes as governance constraints; these are documented governance choices that participants can evaluate when deciding whether to join or interact in a data space.

DIDs and associated mechanisms are the basis for sharing attributes describing a participant and providing evidence about the claims of those attributes in a data space. The participants claims 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.

Instead on relying on a central party to manage authentication and authorisation participants are responsible to manage their own information and to provide this information to other parties in a standardised way.

Every participant is free to choose whether to trust the claims provided by another participant. The core of data spaces is about the mechanisms that enable the verification of participant claims to foster trust.

A core functional mechanism of a data space is the agreement on a common set of processes, reference and rules within a community of participants. The Data Space Governance Authority (DSGA) is a crucial function to support the management of the governance framework of the Data Space by establishing the policies and rules of the data space.

When evaluating different data space architectures and deployment models, the individual set of rules that serve as the basis is important. 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.

Legal requirements might lead to necessary compromises on the goal of participant autonomy and agency. They need to be evaluated carefully to reduce negative impacts on the participants.

13 Foundational concepts of a data space

The mandatory foundational concepts and functional areas of a data space are:

- **Establishing trust:** Defining the procedures, evidence categories, and decision rules that enable participants to evaluate the suitability of counterparties.
- **Data sharing and usage:** Specifying the permitted data flows, processing activities, and contractual obligations that govern how data may be used after it is shared.
- **Data discoverability:** Providing metadata and discovery mechanisms that enable participants to find and assess available data assets and services.
- **Sharing contract negotiation:** Establishing the processes and protocols—ranging from automated negotiation to manual escalation—used to reach mutually acceptable contract terms and policy alignments.
- **Decentralized identity capabilities:** Ensuring participants can represent identity and attributes in a verifiable and privacy-preserving manner; DIDs and verifiable credentials are the recommended approach, while functionally equivalent implementations may be acceptable if they preserve participant autonomy and interoperability.
- **Observability:** Ensuring auditability and runtime monitoring capabilities to detect violations, gather evidence for disputes, and support continuous trust evaluation.

Additional value-adding services that support these main functions of a data space may include the following optional functional areas:

- **Vocabularies and semantic models:** Curated ontologies and schema registries that improve interoperability and precise meaning of data shared across participants.
- **Application and processing services:** Hosted or on-demand processing capabilities (for example, analytics, enrichment pipelines, or confined compute) that offer contract negotiations for API endpoints and enforceable usage policies.
- **Marketplaces:** Commercial catalogues and negotiation platforms that facilitate discovery, pricing, and contractual arrangements for data products and services while respecting governance and access policies.
- **Data trustees and escrow services:** Neutral service providers that hold, mediate, or process data under predefined governance constraints (for example, confined compute or escrowed storage) to enable joint analysis while preserving confidentiality.
- **Other optional value-added services:** Additional capabilities such as notary, or auditing services, and specialised domain-specific tooling which may be selected to meet particular participant requirements.
- **Semantic Models:** Semantic models are listed as optional because their required depth and degree of harmonisation are context-dependent and should be determined by the DSGA and participant needs. In many data spaces, a shared core vocabulary will still be practically necessary for interoperability, even where complete domain harmonisation is not mandated.

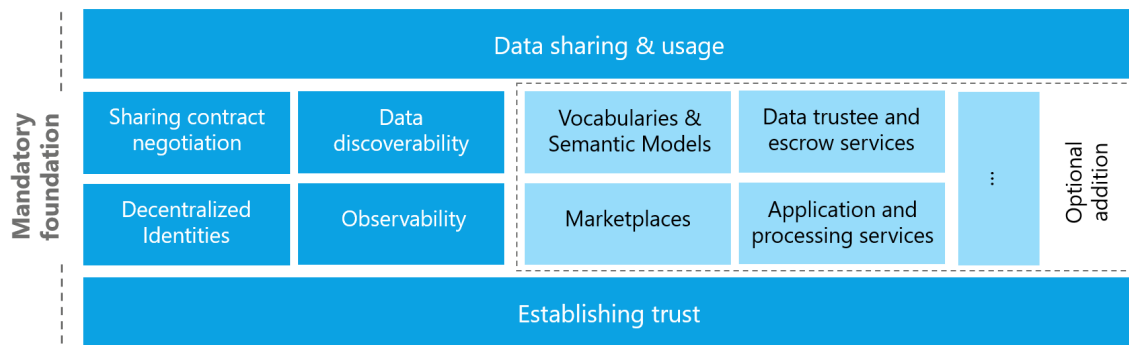


Figure 4: Foundational Concepts in data spaces

14 Establishing trust

Humans build trust with each other by evaluating 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 rules (personal, internal, implicit and/or explicit). 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.

14.1 Establishing trust is the fundamental reason for data spaces to exist

To create value, data needs to interact with other data and then support decision making to enable actions that will create value. The potential to create value increases if data is more diverse, which often requires that multiple actors need to come together and share their data with each other. However, like in any human relationship: Before sharing comes trust. Without trust, the risk of something going wrong seems too high and unmanageable. Creating trust reduces risk. Reduced risk lowers the barrier for sharing data.

14.2 Increasing trust lowers risk

Data spaces can create context-specific trust where trust did not exist before or where it is difficult to establish – for example between competitors. Therefore data spaces reduce the risk of sharing data and through that enable the creation of value.

14.3 Attribute-based trust

Attribute-based trust is a way to establish trustworthiness between two parties based on specific attributes rather than a fixed identity. It functions as a control mechanism with minimal disclosure - proving that certain conditions are met without revealing full details on every interaction.

It's mechanism can be summarised in a couple of simple statements:

- A Participant holds a collection of attributes
- A data sharing contract is a collection of policies
- When negotiating a data sharing contract attributes are matched to policies

A participant's trustworthiness is determined by evaluating their participant attributes. This evaluates the potential risk of sharing data with another participant. Lower risk means higher trust.

In addition to the attributes of the participant the trust level is also based on additional context: the attributes of the data space and the attributes of the data shared in the data space, the applicable trust anchors and trust frameworks, and potentially others.

It can be expressed by complex rule sets that can evaluate many attributes and understand their provenance and who is providing guarantees about them. There is no limit to the attributes that can be defined and the expression of policy rules to evaluate those attributes. It is also possible to define policies that branch into additional workflows, e.g. human approvals, for evaluating claims about the participant attributes.

Depending on the level of tolerable risk (and thus required trustworthiness) 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 cryptographically signed claims, as well as additional attributes that might be provided by the participant directly or indirectly represented through the claims presentation process during the trust negotiation (e.g., proof that commercial contract for the data exists and that payment for the data has been submitted, or proof of technical capabilities: encryption at rest, secure communication channels, etc.).

Attributes can be expressions representing a single claim (e.g., membership credentials of an association) or a set of multiple claims (e.g., the other entity is under a specific jurisdiction and the destination for the data transfer in a specific country). Claims can represent static values (e.g., jurisdiction = country) or contain statements about proofs of technical capabilities (e.g., support a specific encryption algorithm).

Many situations require attributes that are complex and may involve additional workflows, including human intervention. The IDSA Rulebook cannot prescribe a general approach for handling extended, composite, or complex attributes. Their management depends on the specific design and governance rules of the respective data space. Further guidance can be found in the IDS RAM.

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 customised rules. It allows participants flexibility to build and use different implementations based on their requirements.

It eliminates the need for an identity provider that controls absolute decisions and thus removes a single point of control and potential failure. It enables autonomy, agency and thus digital sovereignty of the individual participant.

15 Attributes & Claims

As described in the section Establishing Trust Attributes are expressed by Claims that contain Evidence about those Attributes. Matching them to Policies and verifying that the desired conditions have been met creates trust.

Attributes describe a participant and are presented to other participants as claims. A **claim** is a **credential**, which must be verified by cryptographic means that it has been issued by a specific party. This issuing party thus acts as a root of trust and is a **Trust Anchor** for this claim. **If a verified claim and its trust anchor are accepted, evidence about the attribute has been established.**

Each credential must be verifiable, which means it needs to be rooted in one (or more) trust anchor(s). This is a mechanisms that citizens of every country use in their daily lives: Trust depends on the authority that issues them, such as a department of traffic issuing drivers licenses or a government issueing citizen ID cards or passports. The driver license contains the information, like date of birth and name of the holder, but what makes this information trustworthy is the believe that it is not a fake id but one issued by a trusted organisation, the department of traffic. Digital credentials have a big advantage versus the analog examples used here: it can be cryptographically verified whether they are real or fake.

Using a trust anchor does not imply dependence on a single central authority. Participants, together with the DSGA and the applicable DTFs, define which trust anchors are acceptable. Multiple anchors may be accepted to preserve decentralization and choice; the DTF and DSGA should document acceptable anchors and any validation procedures.

A trust anchor is an entity that issues certifications about an attribute. Dataspace Trust Frameworks (DTFs) often contain a list of credential issuers which are accepted as trust anchors for that DTF.

Trust anchors are well known by the DTF, DSGA or Participant accepting them for their capability to verify attributes and for being trustworthy that they enforce compliance with the party that they issued a credential for. 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. Therefore the government of this country can act as the trust anchor for company registration information.

Deciding which trust anchors, and thus which rulesand procedures of issuing and validating attributes are used, is the responsibility of each participant and can be greatly simplified by the DSGA and the DTFs.

In order to use of the concepts described above, every individual data sharing contract needs to provide information about which trust anchors are accepted as roots of trust. This can be done by pointing to DTFs as a shortcut for multiple policies. See the section on policies for more details.

15.1 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 to be able to provide a self-description that follows a known format and protocol, as well as a semantic model that describes the meaning, naming and formt of the attributes.

The format of the self-description can be defined through the onboarding process for a specific data space and may be a part of the membership policies. In many cases, the format and semantic model of the self-description also depend on the selected trust anchors and DTF.

A data space might even require multiple self-descriptions (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 rules provided by the DSGA.

The technical representation and communication of the self-description thus may vary from one data space to another and will be influenced or mandated by the DSGA.

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 is used in many policy evaluations throughout the data space. A non-exhaustive list of examples is:

- Information for the onboarding 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 software agent with this participant, what encryption techniques are supported. Whether this information is stored and distributed in the same way as claims about attributes is a question of the data space requirements and tools used.

16 Policies

Policies are the rules by which trust boundaries are established within a data space. They are used at multiple levels and at almost every interaction point. The main policy groups that are central to the functionality of a data space are access policies (which control access to something) and contract policies (which control specific terms of a contract). 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 the 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.

Ultimately, the sum of all policies (data space, participant, asset) form the trust context on a specific data sharing contract governing the access and the use of a specific asset.

Therefore, data spaces and participants 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.

Dataspace Trust Frameworks (DTFs) can provide policy building blocks to simplify the design and selection of effective policies to reach the desired trust level.

16.1 Policy Design

Policies generally express three classes of constraints used to control interactions:

- **Prohibitions:** Explicit assertions of forbidden actions (for example, “data may not be exported outside the EEA”), which must be enforced and may be paired with monitoring and sanctions.
- **Obligations:** Required actions or checks that a participant must perform (for example, logging, periodic revalidation, or notification), often with accompanying deadlines and verifiable evidence.
- **Permissions:** Explicitly allowed actions that define the scope of permitted processing or sharing (for example, “aggregate-only analytics allowed”), which define the positive authorisation space for consumers.

Constraints expressing a rule can be combined into more complex rules, which then form the applicable policy. For example, a data space participant may only allow access to certain 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 further processing if they meet a set of security standards.

16.1.1 Membership Policies

As discussed above, the first line of policy constraints are the membership policies and processes required to join a data space. These policies ensure that only organisations with specific attributes they can verifiably prove, can join. These can be policies that verify the applicant’s HQ location, industry certification, membership in industry associations, but also policies that would require human interactions and complex workflows, such as

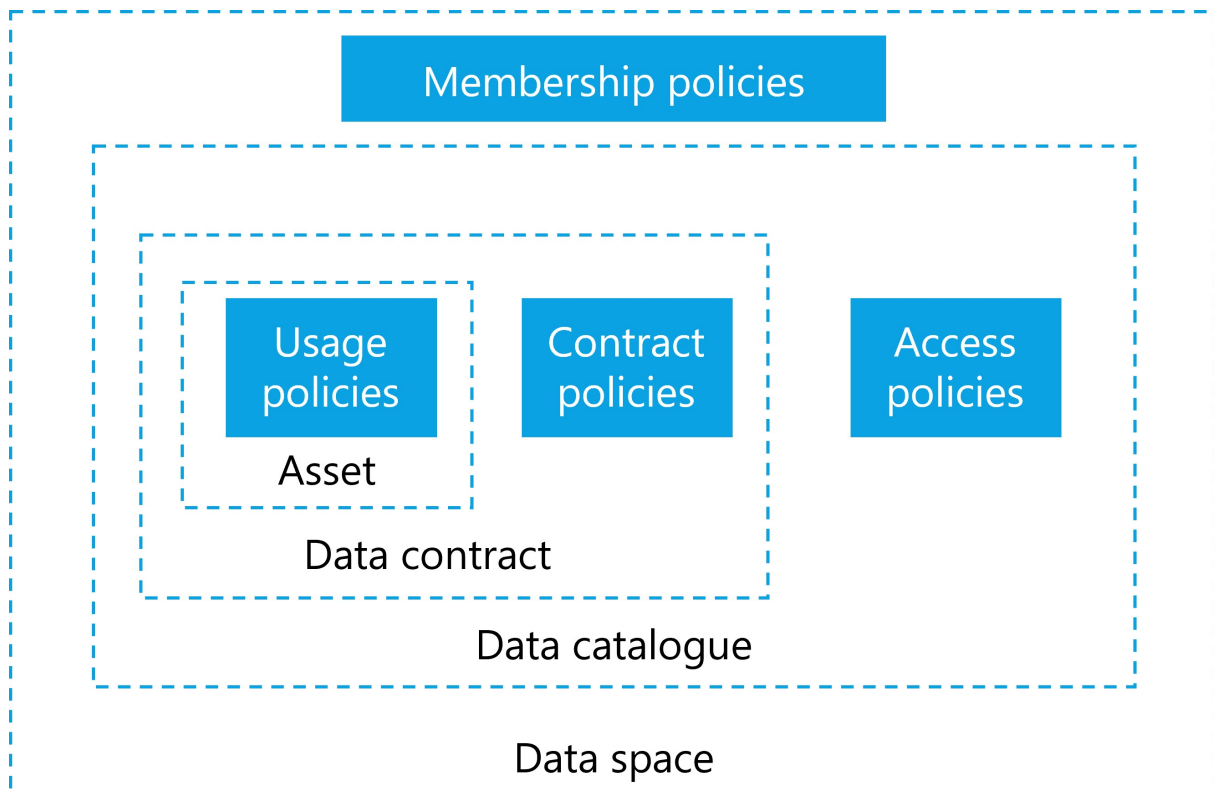


Figure 5: Different policies in data spaces

a valid contract with an organisation implementing the DSGA that must be negotiated before an applicant can become a participant.

16.1.2 Access policies for data discovery

Once an applicant becomes a participant, the next layer of policies becomes relevant: access policies. An access policy defines which attributes must be available to discover and access data contracts within a data catalog. A participant that does not have access to a specific data contract should also not be able to discover the contract offer in the catalog.

Optional services, like a marketplace, must adhere to this principle as well and only show items based on matching access policies and participant attributes.

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).

In a scenario where contract offers should be made visible to everyone (even non-participants), the access policy can also be expressed as an empty policy, not triggering any restrictions and thus making the contract visible to everyone (e.g. Open Data).

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

16.1.3 Contract Policies

When a participant has access to a data contract offer the next set of policies comes into play. A data contract offer can have contract policies that define what attributes are needed for a data contract agreement. Contract policies 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. A more complex attribute example involving human interaction is the association of the data contract with a legal contract between the two parties, which typically occurs outside of the data space technology and processes. The negotiation of policies can be anywhere on the spectrum of 100% machine-processable, automatable and immediate to a human workflow potentially taking a long time.

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

16.1.4 Usage Policies

Contract policies may also include usage policies that take effect when the data is shared and control how the data can be used by the consumer. Depending on the value of the data, use cases, trust levels, contracts in place and many more attributes, there are different possibilities to enforce usage policies which come at varying costs.

For data with low value or data not under a specific legal protection, it is too expensive to build a system that guarantees control. It is 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 of high value and require stronger protection and higher technical costs.

When designing a data space and deciding which data to share, it is important to understand data 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 without incurring unnecessary cost or constructing detrimental barriers to sharing.

Enforcement guidance (example mapping):

- **Low sensitivity:** Monitoring and contractual remedies; participants are expected to perform periodic verification and rely on legal enforcement when misuse is detected.
- **Medium sensitivity:** Robust logging, decentralized observability, periodic automated checks, and contractual remediation; independent observers or auditors can be used for dispute resolution.
- **High sensitivity:** Strong technical enforcement (e.g., confined compute, mandatory encryption, short trust validity), on-demand revalidation of claims, and involvement of observers/auditors as participants. Real-time or per-execution checks are recommended.

Who monitors: enforcement and verification are primarily the responsibility of participants. Decentralized observability services, auditor participants, or notarization services can be used to provide independent evidence; the DSGA should define expected verification intervals and monitoring roles for each sensitivity class.

Example	ProtectionNeed	Explanation
Public weather data	low	Some data sets are already publicly available and can be shared with low friction.
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, transactions and be misused (e.g. for investment purpose).
Personal health data	high	Personal health data are highly protection worthy, regulated by strong laws and pose potential danger to the individual in case of data misuse.
Machine operation 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.

16.1.5 Policies for segmentation

Policies can also act as a fine filter to segment data sharing scenarios and use cases. Adding a single policy that checks for the membership credential of a distinct group is the equivalent to restricting this specific contract to that closed group of organisations. Segmentation through policies can be used by an organisation to easily manage the sharing context of their data assets and can be applied at many different layers and segments. Here are some examples of what segmentation can be achieved with policies:

- **Data space boundary:** Enforce segmentation by membership credentials, ensuring only participants who possess a valid membership credential for the data space can access protected resources.
- **Business use case:** Restrict access to contracts or datasets by requiring credentials or attributes that represent participation in a particular business use case or consortium.
- **Individual participant filtering:** Apply policies that target specific participants by identity attributes or credentials (for example, to whitelist a participant).
- **Technology compatibility:** Limit interactions to software agents or endpoints that meet defined technical interoperability or security profiles, ensuring compatible communications.
- **Geographic constraints:** Require credentials or attributes indicating jurisdictional compliance, preventing access where legal or regulatory restrictions apply.

- **Industry association or certification requirements:** Enforce access based on membership of an industry association or possession of required security certifications.

... and many more. The possibilities to enforce control and segmentation are unbounded.

17 Data space membership

Technically the membership in a data space is just a credential. Nothing more than that. If a data space software agent can show the membership credential it acts on behalf of an organisation which is a member of the data space.

However, on the business and legal layer things can get more complicated. There the membership in a data space can be a complex set of business process that need to be followed and legal documents that need to be signed.

To be able to issue membership credentials a data space needs to define policies that specify what attributes an applicant must meet to become a trusted participant, often also referred to as a member. The data space governance authority (DSGA) manages those policies, and it defines the business processes that might be necessary for this onboarding. As policies need to be clearly understandable and interpretable the DSGA also needs to provide a semantic model that details the defined policies so that applicants can correctly interpret the policies and provide proper claims about their attributes. The referenced semantic model describes the acceptable policies, their names, the potential value, and the format in which those values are accepted.

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

18.1 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.

Note: The disadvantages listed for decentralized identities describe governance and operational trade-offs, not a loss of participant agency. DSGAs may define accepted trust anchors and onboarding processes; participants remain free to accept or decline those trust anchors and thereby preserve agency in their decision to join or interact.

19 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 a description of the data space that specifies its participation requirements. 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, machine-readable discovery protocols of existing participants among other things.
2. Candidate reads the requirements 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 onboarding services defined by the DSGA (for example, one or more onboarding credential issuer services). The technical implementation may vary based on the requirements.
5. The onboarding services request 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 onboarding service 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 assets of others, negotiating data contracts).

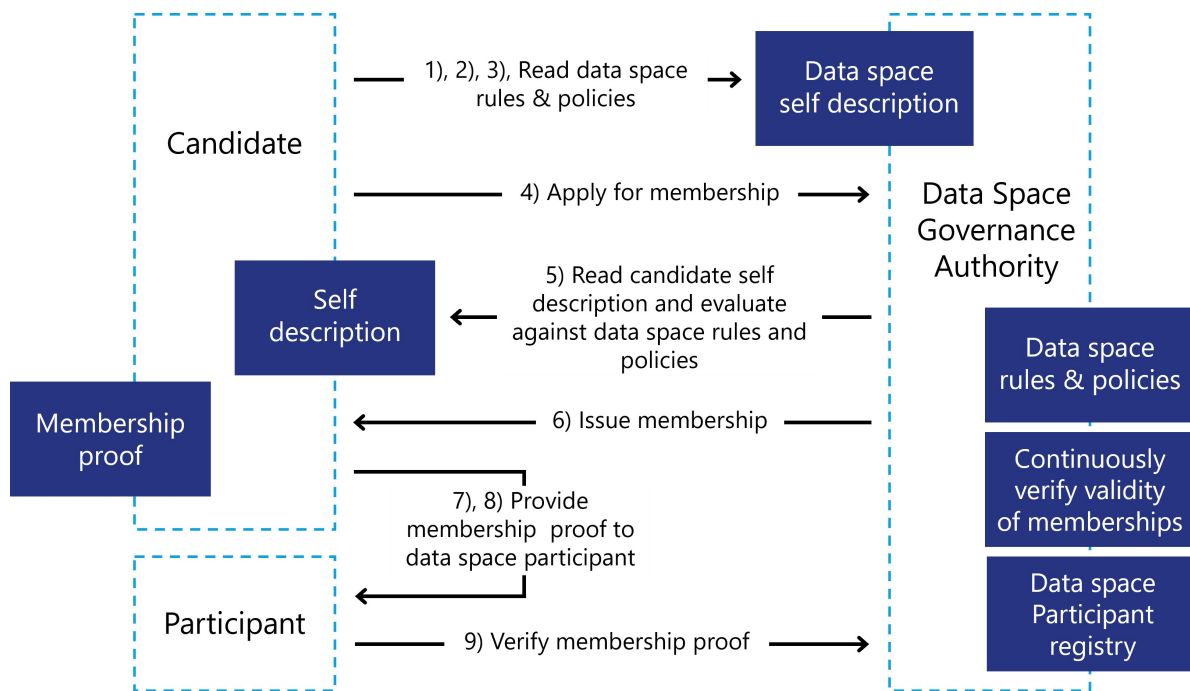


Figure 6: Onboarding in data spaces

20 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.

20.1 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 Data Contract Offer (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).

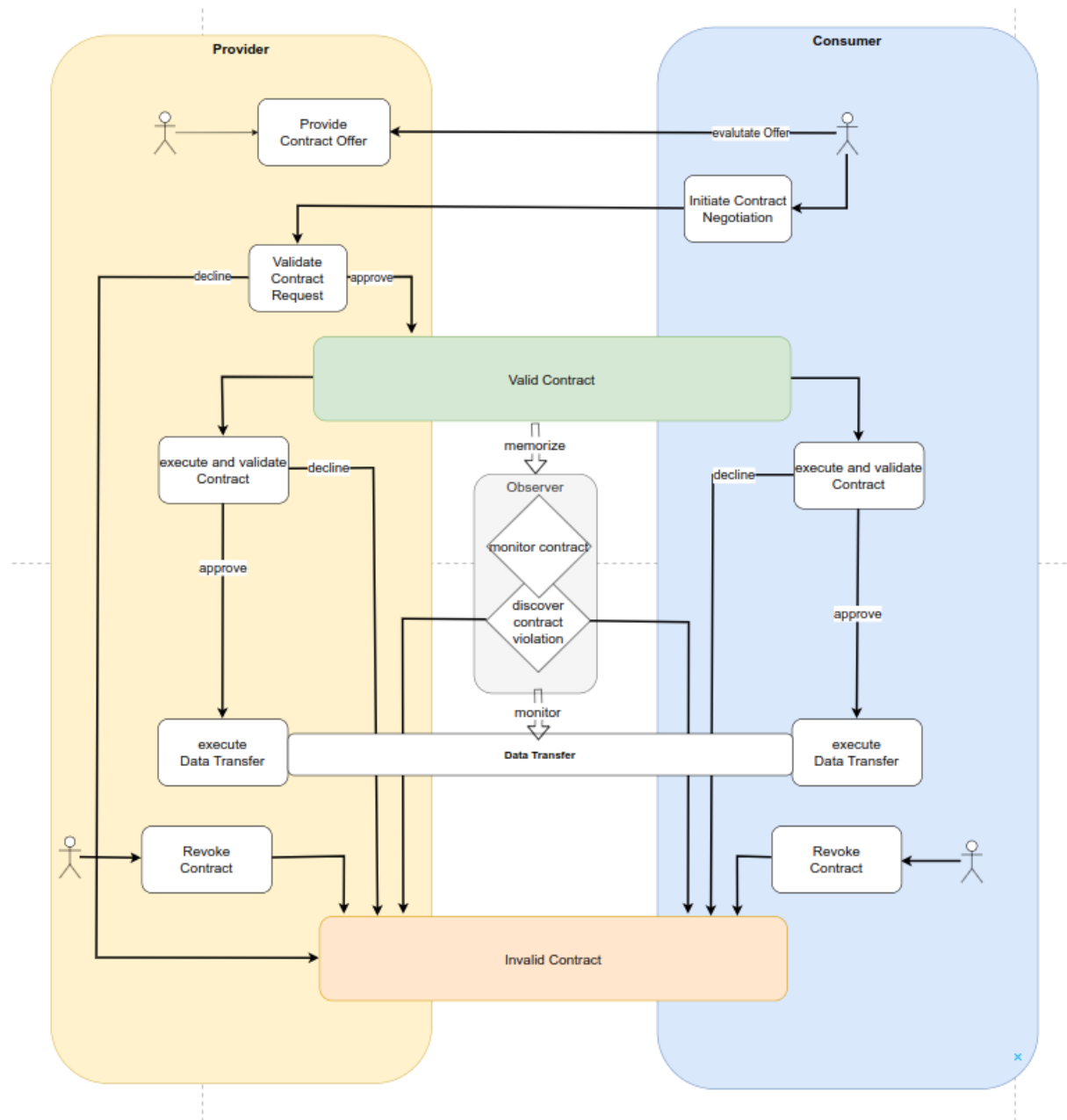


Figure 7: Data sharing contract negotiation

20.1.1 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 software agent 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 centres to enable a highly available data sharing architecture.

When data is not moved but a “code to data” approach is selected, the push and pull behaviour 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.

21 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.

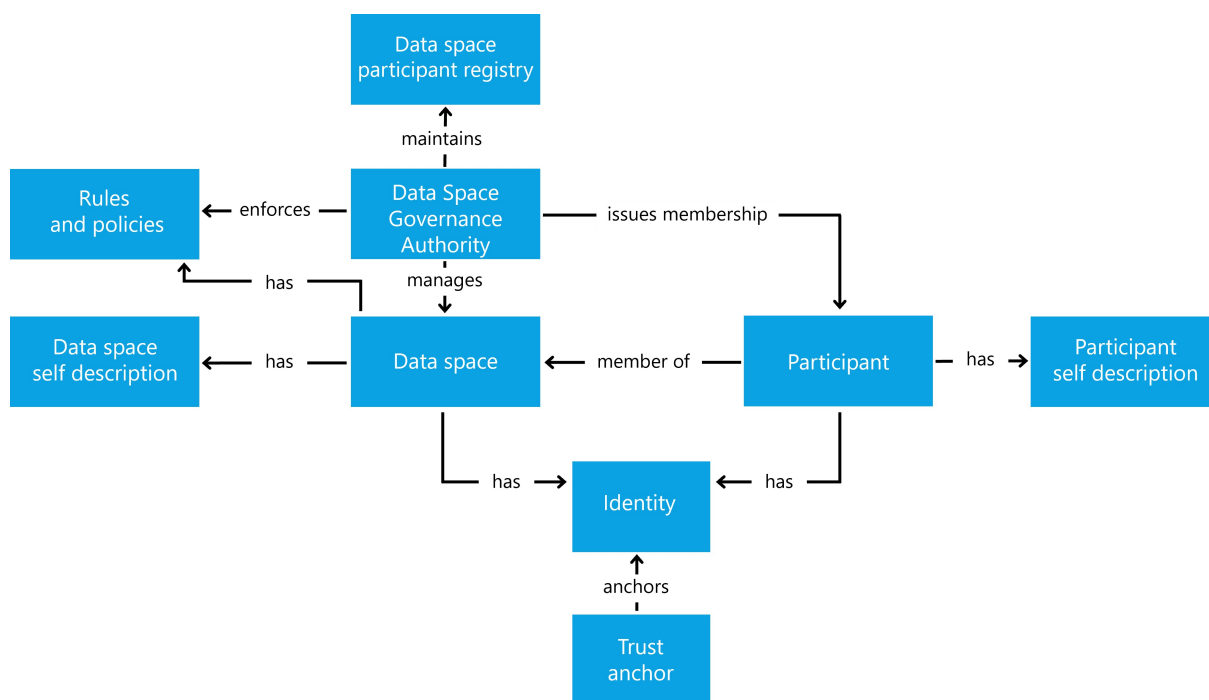


Figure 8: Overview of Data Space entities

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 behaviour) 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 following functional elements must be specified and documented:

- **Rules:** Define the expected behaviours, roles, and minimum technical and organisational capabilities required of participants (for example, required legal contracts, organisational certifications, staff skills, security practices, or operational governance targets).
- **Policies:** Specify the participation, access, contract, and usage policies that will be enforced by the data space participants and describe how these policies are verified and evolved.
- **Membership certification:** Describe the mechanism and evidence required to verify and certify membership (for example, identity documentation, audits, or third-party attestations), including issuance, renewal, and revocation processes.
- **Participant discovery and self-description:** Establish if, where, and how authorised participants and their self-descriptions are published or discoverable, and which attributes are visible to different query roles.
- **Identity system:** Decide whether identities will be managed via decentralized identifiers (DIDs), centralized identity services (not recommended), or a hybrid approach, and document the control, governance, and privacy implications of each option.
- **Catalog(s):** Specify the catalog model (single central catalog (not recommended), multiple federated catalogs, or per-participant catalogs), their APIs, metadata standards, and how catalog visibility and access controls will be applied.

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

The self-description for the data space should include the following documented elements:

- **Membership policies:** The eligibility criteria and evidence required for organisations to join, including any onboarding checks, audit requirements, and renewal conditions.
- **Trust anchors and trust frameworks:** A list of accepted credential issuers, Dataspace Trust Frameworks (DTFs), and the rules by which claims and attestations are validated.
- **Attributes used for trust decisions:** A clear listing of attributes, their meaning, and how they map to policy constraints so participants can determine how trust levels are calculated.
- **Technical component requirements:** The required interfaces, protocols, and implementations (for example, software agent capabilities) necessary to interact with the data space.
- **Participant discovery and self-description endpoints:** If participant discovery is required, define the location and format of discovery endpoints and self-description publication, and which attributes are public versus private and only accessible to existing members.
- **Registration service and related workflows:** The documented workflow to request membership, the validation procedures used to determine compliance

with membership requirements, the process for issuance of membership credentials, and the rules and procedures to revoke credentials when necessary.

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

Once at least one Onboarding Service implementing the DSGA is instantiated, organisations can apply for membership. The functional elements listed above are mandatory capabilities for a functioning data space, but they may be implemented in decentralized or centralized ways. When a centralized or federated implementation is chosen, the DSGA must document the reasons for that choice and describe mitigating measures to protect participant autonomy and agency. 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.

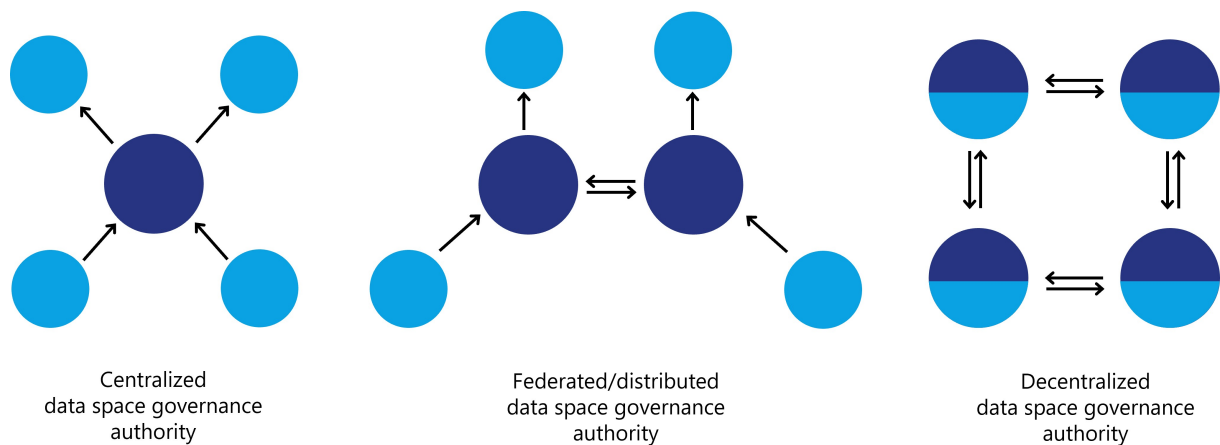


Figure 9: Variants for data space governance authorities

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

22.1 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 organisations 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, organisational, 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.
- Organisational 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.

22.2 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 inter-

operability. 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 organisational 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 organisation 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 organisation 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 organisation of the Data Space, it is also possible that a DSGA exists without any legal organisation 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 organisation 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.

22.3 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.

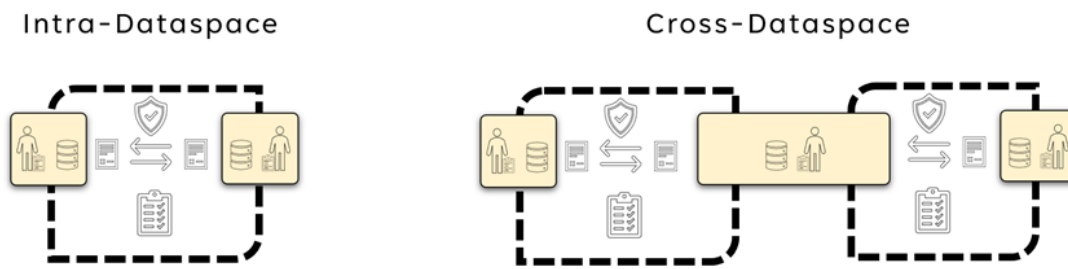


Figure 10: Interoperability Models

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.

22.4 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 harmonised 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:

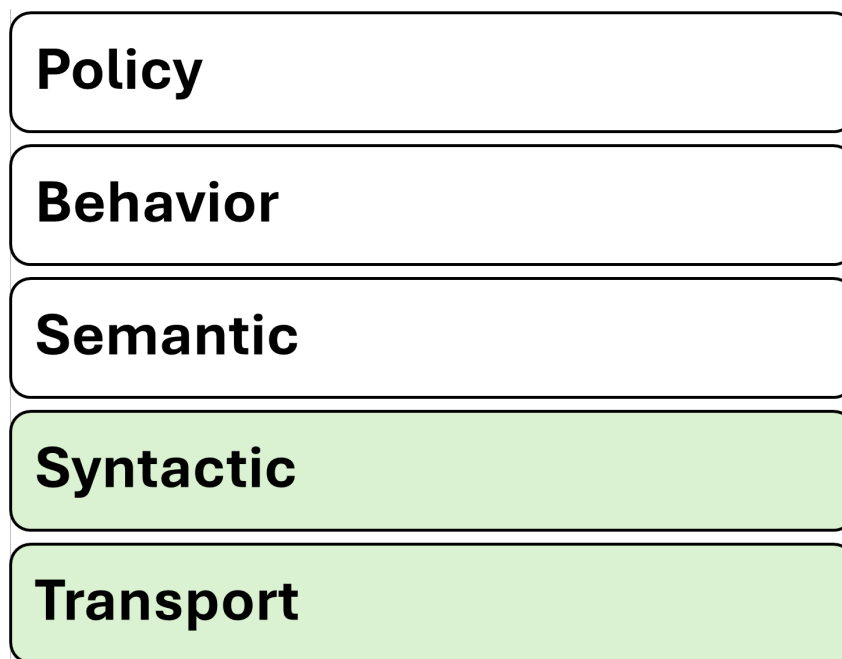


Figure 11: ISO 19941 - Cloud Computing Interoperability and Portability

And second the European Interoperability Framework facets:

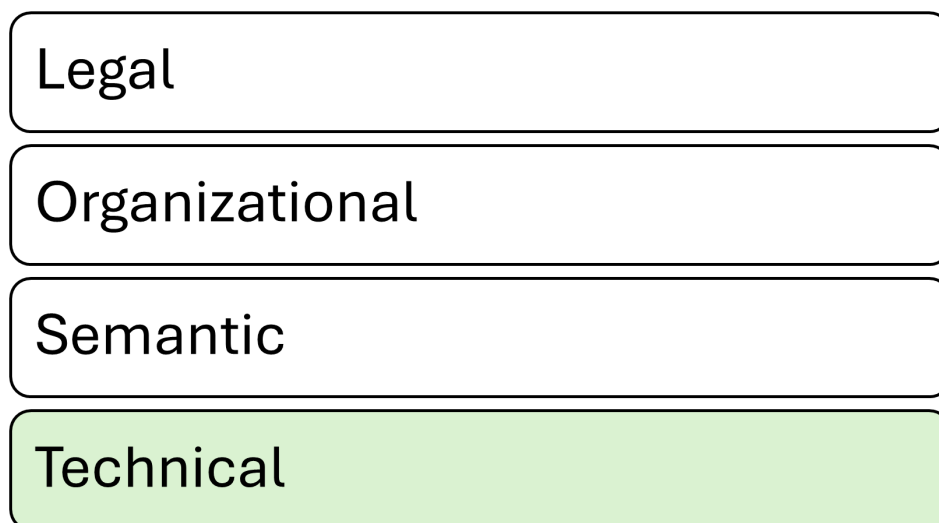


Figure 12: European Interoperability Framework

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.

22.5 Interoperability facets in Data Spaces

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

22.5.1 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.

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.

22.5.2 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.

22.5.3 Organisational

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

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

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

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

22.6 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, utilising 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 organisational 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.

22.7 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.

22.8 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 utilising 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.

Recognising 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 organisational 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 organisations.
 - **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 organisational 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 standardised 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.

23 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 in Creating a Data Space). 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.

23.1 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 Central control – a gatekeeper can regulate which entries are permissible and which are not Easy discovery as only one catalog needs to be queried	A central gatekeeper can arbitrarily exclude participants and their data from the catalog Single point of failure Potential performance bottle neck Security issues will affect all members at once

Catalog architecture	Advantages	Disadvantages
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
Decentralized catalog	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
	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 provided through the DSGA or discoverable through a peer-to-peer protocol. The DSGA should specify the chosen catalog architecture and justify any centralized or federated choices, documenting the associated trade-offs and the mitigations used to preserve participant autonomy and neutrality.
	Data Space as a whole is more resilient towards cyberattacks even though individual members can experience outages	Participants need to crawl each other's catalogs to see which items are available
	Easier to scale	

23.1.1 Access policies

A best practice of access security is for an IT system to show users only what they need to know - to minimise 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 authorised to request a contract negotiation. This does not imply that the participant already has authorisation 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.

24 Observability

This chapter aligns with the IDSA position paper *Observability in Data Spaces* and uses the terms *observability* and *observer* as the primary concepts for this functional area.

Many times it is necessary to make the data sharing process observable. This can be done for legal reasons to prove that data has been shared only with authorised entities, or for business reasons such as billing, dispute resolution, and compliance evidence.

To enable observability in practice, related business processes between participants (for example log exchange, evidence retention, dispute handling, and escalation paths) should be defined and, where required, included in the governance framework of the data space.

Depending on the architecture of the data space, multiple solutions are possible.

- Centralized or federated observer model, which has two main shortcomings when implementing large-scale data spaces: It presents an additional vulnerability that could affect the sharing of mission-critical data. Also, a central observer with information on all data sharing contracts represents potentially valuable knowledge about participants. This can be exploited for financial gain, making it a target for bad actors.
- Decentralized architecture can minimise the risks associated with a centralized or federated observer model.

In a decentralized observer architecture, every participant keeps the information about the agreed contracts 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. This log data can be shared with a participant that is providing the observer role. The observer then matches the corresponding logging information and reports any irregularities to the parties participating in the sharing contract (or to the respective regulator if required). The same mechanism can also be used for example for billing and notary functions.

The observer role may be assumed by the participants involved in a data sharing transaction or by third-party participants identifiable through specific credentials (for example an industry auditor rooted in a governmental trust anchor).

E.g.: If regulation requires an auditor to review data sharing information this information can be obtained through dataspace mechanisms. To audit the contracts of a participant, the auditor requests the log data which is published as data contract offer with an access policy restricting access to the auditor. To verify the validity of those log entries, digital signing mechanism are used or the corresponding log data from other participants can be requested to reconcile claims. This limits 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 observability 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 observability in a data space, the DSGA can mandate that participants make their observability data available through data sharing contracts for ongoing event notifications or streams by default.

Following the same pattern, additional optional functional roles can be implemented: invoicing, billing and payment clearance services, notary services, regulator reporting, and

the like.

Reference: Observability in Data Spaces (IDSA Position Paper)

25 Vocabularies

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

- **Semantic models for policies:** Structured models that define the terms, attribute names, and allowed values used in policy expressions; these models make policy requirements machine-interpretable and reduce ambiguity during reconciliation.
- **Semantic models of the shared data assets:** Domain-specific data models and ontologies that define the meaning and structure of shared datasets, enabling consumers to interpret and process data correctly and to enforce usage policies based on data semantics.

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.
- Within the Data contract as a machine-readable specification that describes the contractual terms, associated policy constraints, and the semantic models required to interpret the asset. The data model, vocabulary, or usage rules that must be understood to correctly interpret and enforce contract-specific policies (for example, an industry-specific usage restriction expressed as a domain ontology).

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.

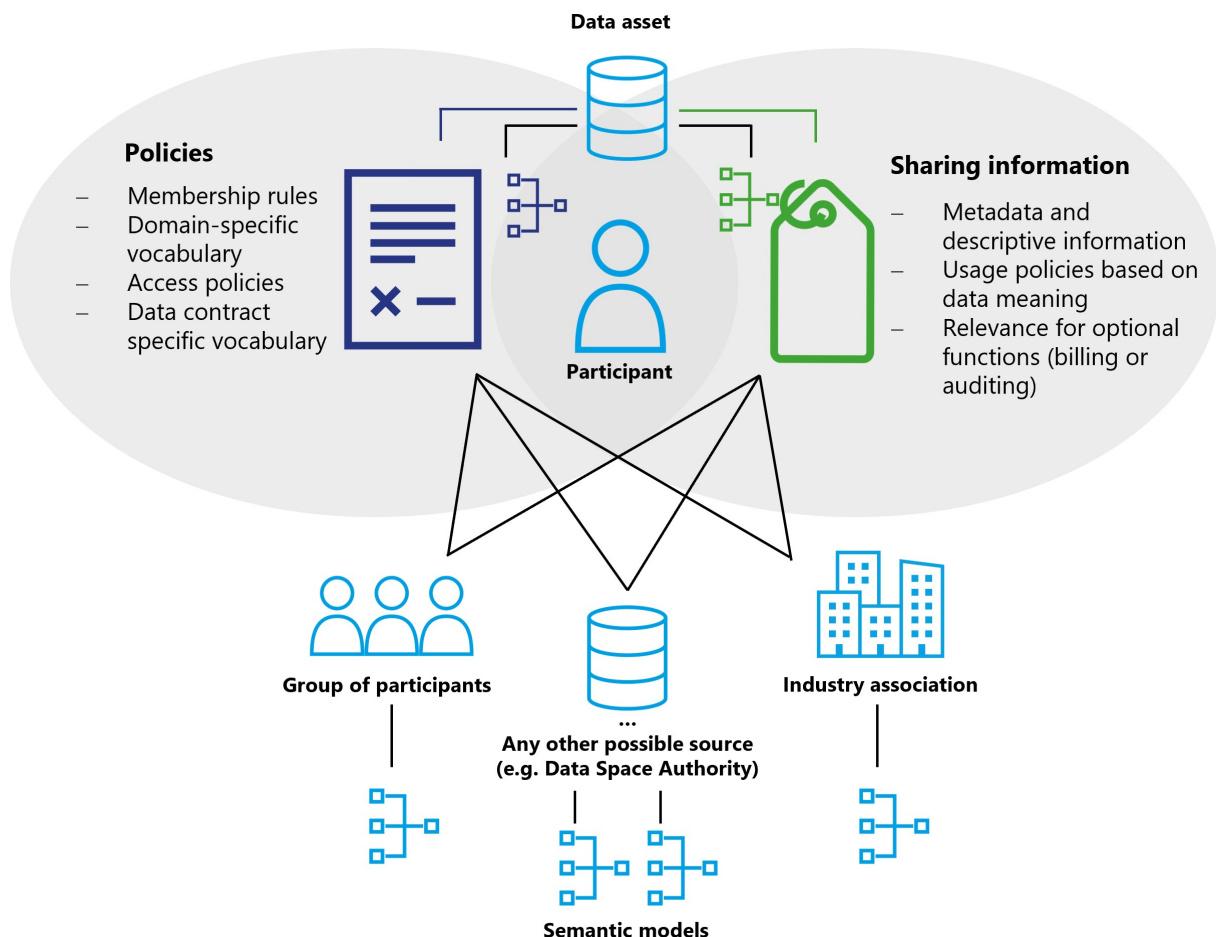


Figure 13: Vocabularies and their relationship to data assets

25.1 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 fulfil all requirements, policies and procedures enforced by the DSGA to establish trust.

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 3.0 in more detail. The IDSA Rulebook provides a limited explanation below.

26 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 dataspace 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.

27 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.

27.1 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.

28 Connector

A connector is a specific implementation of a software agent for accessing data spaces. It 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.

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

29.1 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 organisations 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.

29.2 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!

29.3 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 its 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.

29.4 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 organisation that needs to join a dataspace or use dataspace protocols and processes to access and use data?

29.4.1 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 organisation 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.

29.4.2 Agent First

Agents discover data offerings that other organisations are publishing through dataspace but don't yet have existing access. Depending on the rules of those dataspace 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 organisation to manually join the dataspace. To negotiate access ad hoc, Agents can leverage dataspace protocols using decentralized claims and policy mapping for dynamic, cross-organisational attribute access control to be granted permission to shared data and to agree on data usage contracts.

29.5 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. Dataspace 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:** Standardises 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:** Standardises interactions between organisations. 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 DCP 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 Organisation A needs data from Organisation B the two organisations 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 organisations can be established when one organisation defines policies for trustworthiness and the other organisation provides evidence—through verifiable claims—that these policies have been met. By verifying these claims, both organisations 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 Organisation 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 fulfil 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.

29.6 Semantic Interoperability and Compliance

As it will be impossible to create a globally standardized semantic model for data sharing contracts AI can help with the interpretation of different models and therefore aid in the fulfillment of the data sharing contract. AI's ability to map different semantic models reduces 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 integration.

29.7 Strategic Recommendations

1. **Invest in Protocol Standardisation:** Support the development and adoption of MCP, A2A, DSP and DCP to enable secure, scalable AI integration.
2. **Prioritise 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.

29.8 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. Organisations that embrace these innovations will lead to digital trust, compliance, and business agility.

30 Onboarding Pattern for Decentralized Data spaces

30.1 Joining a data space

Decentralized data space architecture emphasizes the autonomy and agency of participants, thus it must be a voluntary decision of a participant to join a data space.

Decentralized Data Space Architecture

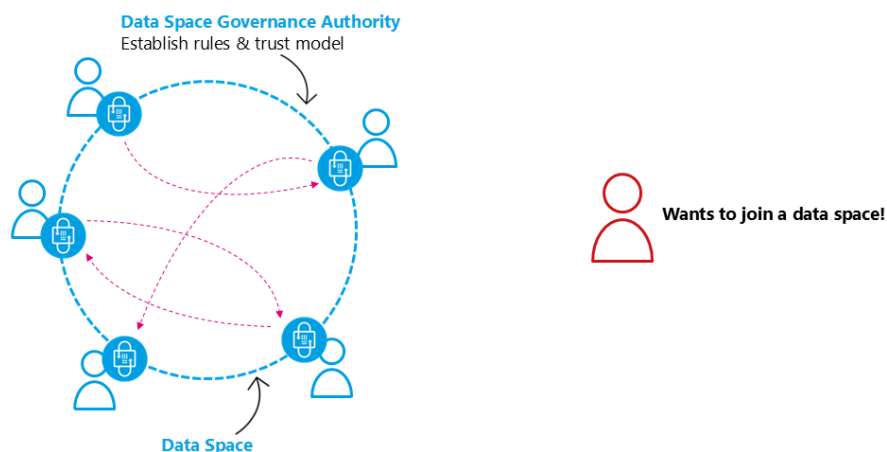


Figure 14: a new organisation wants to join a decentralized data space

30.2 Selecting a hosting model and provider

However, with great power also comes great responsibility, so the participant has some work to do. It's not like joining a centralized platform, where one creates an account and that's it. The participant needs to actively make a decision on where to operate the technology that's needed to be active in a data space: The data space software agent (e.g.: connector), the wallet, data staging areas, etc... Although this seems to be a complex and difficult task to accomplish it is ultimately what enables digital sovereignty. Every participant can operate on the infrastructure of their choice and they can make an autonomous decision on how much agency to have. Any operating model available - from on premises to classical cloud computing is also possible for data space technologies: starting at a fully customised hosting model on one end of the spectrum, a platform as a service model for software agents (connectors) and wallets, all the way to fully managed use case services (e.g. a data intermediary offering a Digital Product Passport service including sharing capabilities to a manufacturing data space) on the other side of the spectrum. Further, such components might be operated for an individual data space or reusable for a multitude of data spaces by leveraging dynamic contexts and configuration capabilities.

30.3 Onboarding to a data space

Once the future participant has all technical components up and running they need to understand the rules and processes of the data space, maybe sign some legal contracts of the data space organisation, provide evidence on their claims that they satisfy the rules of

Decentralized Data Space Architecture

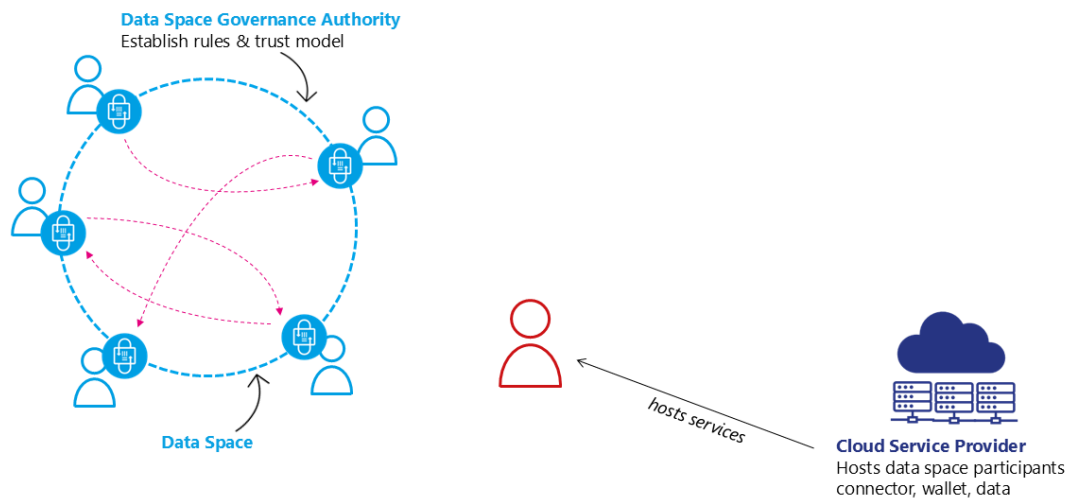


Figure 15: the organisation provisions data space resources with the infrastructure provider of their choice

the data space and many other steps needed to be onboarded to the data space. This can be done by an onboarding service provided by the organisation representing the community of the data space (e.g. a not for profit organisation, a government entity responsible for a regulatory data space, or a commercial entity leading a data space created around their organisation and its partners). This onboarding service is needed only as a one-time access point while joining the data space. It might have an additional service like notifying existing members that a new member is joining, but usually will not be needed for the ongoing operation of the data space.

It is important to understand that in a decentralized architecture, no ongoing dependency on a central or federated services should be mandated. As this example illustrates on popular mechanism to realize a decentralized architecture there are other solution paths as well.

If the data space relies on an onboarding entity to manage the onboarding process this entity can ensure the adherence to membership rules and potential additional steps like checking the validity of external DTF credentials.

If there is no onboarding entity a new participant could simply prove their eligibility to be a member of the data space by providing claims to another participant who accepts the membership after evaluating all membership rules and the claims provided.

There are many other models of how membership of a data space can be established. For further process steps in this article the focus will be on the scenario of an Onboarding Entity working with an external credential issuer.

Often the Onboarding Entity can directly issue membership credentials for the new participant, however, in this scenario it requires additional steps from external actors, the Credential Issuers from a DTF which are being used within the data space.

Decentralized Data Space Architecture

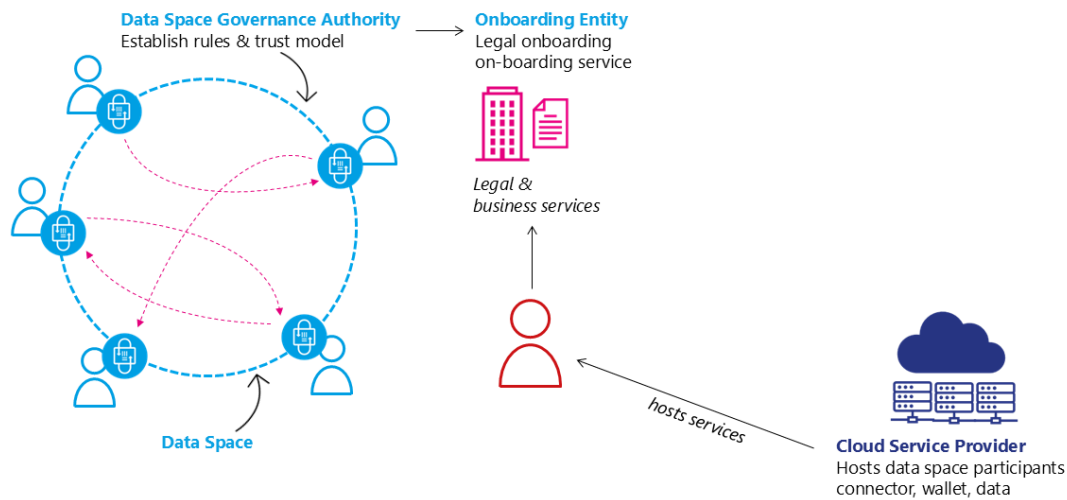


Figure 16: An On-boarding Service verifies compliance with the rules of the data space

30.4 Obtaining Credentials from external data space Trust Frameworks

In complex, large-scale data spaces a data space Governance Authority will not define every single rule of the data space but rely on existing data space Trust Frameworks to provide building blocks from which to construct the DSGAs rules for this specific data space. E.g. Gaia-X could be leveraged to provide labelling for the infrastructure providers used by individual participants, iSHARE could be supporting the legal onboarding of participants. In those cases the Onboarding Service will need to collaborate with the onboarding/issuing services of the various DTFs. However, the ultimate responsibility of maintaining credentials about their claims is with the participant. So although the Onboarding Service of the data space might collaborate with the Issuance Service of the DTF the DTF credential has to be issued directly to the data space participant. Alternatively the credential issuance can also be requested by the participant.

This process might have to be repeated for every credential issuer that is being used within the data space. It might also be possible that the participant already holds a credential from accepted external DTF credential issuers due to their membership in another data space. It is possible that this are accepted by the data space.

Once issued the data space participant can use this credential together with the membership credentials from the data space Onboarding Service to provide evidence of its claims when negotiating a sharing contract with other participants of the data space.

It is important to note that many variations of this process can exist and it is part of the role of the DSGA to define the exact business process which leads to the set of credentials that are then resulting in a membership credential for the data space in question. This process can be influenced by business needs, regulation, and other external factors.

Decentralized Data Space Architecture

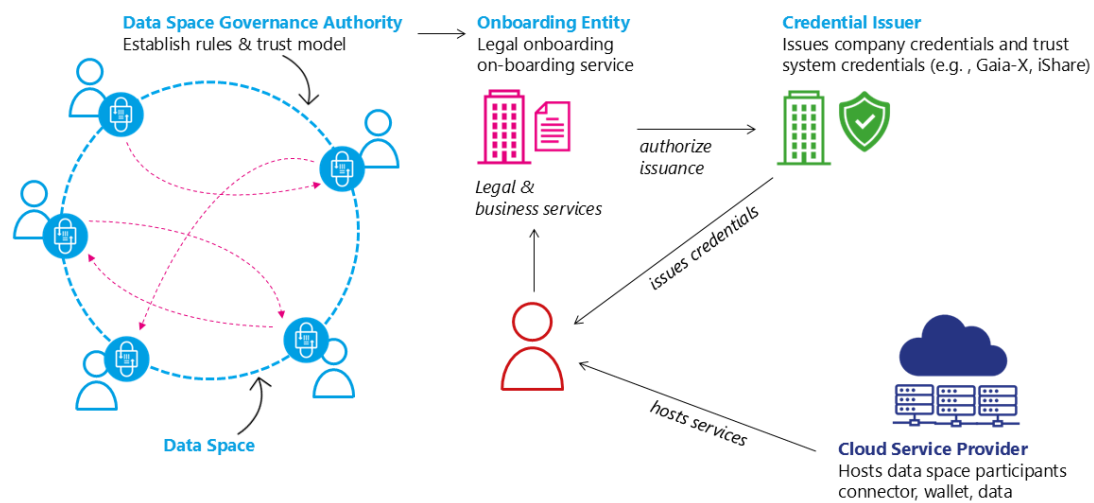


Figure 17: Credential Issuer issues credentials

30.5 Using credentials in the data space

When two participants of a data space interact it will be necessary to match policy constraints of one party to the claims that the other party presents. In most cases it will be sufficient to verify the signature of the credential issuer and the expiration date as the claim is presented by the participant. After all cryptographical methods will enable the other party to verify that the presented credential is representing the counterparty and has been signed by the credential issuer. However, in more complex cases it might be necessary to check the validity of a signature synchronously with the Credential Issuer. One such example can be the necessity of near real-time checks of the revocation of a credential.

Decentralized Data Space Architecture

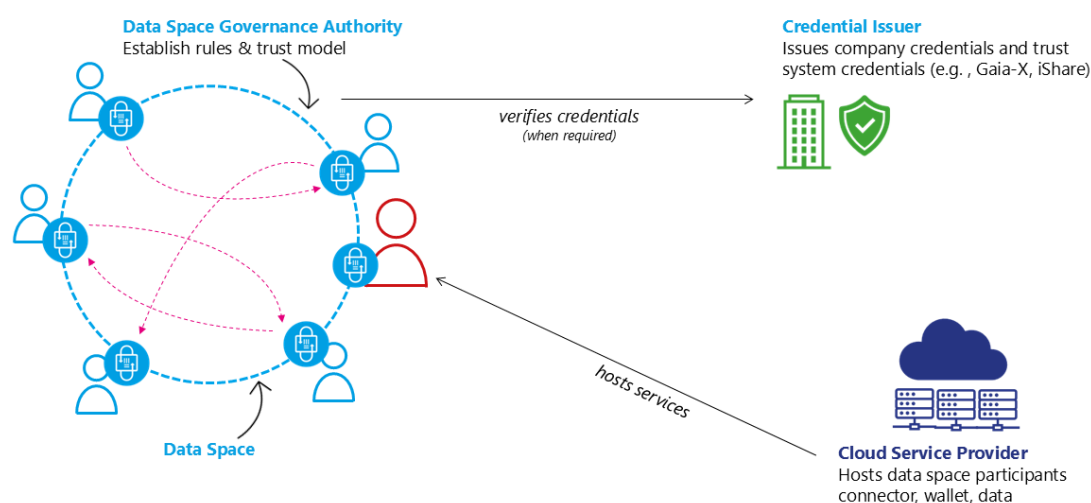


Figure 18: Credentials are verified by other participants

31 Summary and outlook

The IDSA Rulebook recognises 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 plurality the current global regulation and legislation. Data spaces are a durable concept, applicable to data ecosystems internationally.

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 and management of data spaces.

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.

Based on the recent IDSA work, additional publications provide more insights into data spaces. As concept of data spaces evolves updates to the IDSA Rulebook will be made. As many aspects of data spaces are updated separately and the urgency of discussions reflects and ever changing landscape of focus points of attention the IDSA Rulebook has been structured in a way that allows for fine granular update of the presented concepts.

The IDSA Rulebook will no longer feature a fixed version number. The published version always reflects the latest design aspects and will be updated as new contributions are made to the GitHub Repository, discussed in the Working Group Meetings and approved and published. A versioned history of each document is available through the detailed version management mechanisms of GitHub.

CONTACT

International Data Spaces Association

Emil-Figge-Str. 80
44227 Dortmund | Germany

phone: +49 231 70096 501
mail: info@internationaldataspaces.org

WWW.INTERNATIONALDATASPACE.ORG



[international-data-spaces-association/](https://www.linkedin.com/company/international-data-spaces-association/)