



iSHARE

# COOKBOOK FOR DATA SPACES

Quick Start Guide for Data Spaces

Gerard van der Hoeven & iSHARE Team

Version 1.3 | August 1, 2023 | © iSHARE Foundation 2023

# Preface

This book arose from the conviction that it is necessary to explain the thinking about federated data exchange in data spaces in a simple way, but at the same time to provide a starting point for data initiatives to take practical steps.

The world of data spaces is constantly evolving and elements are already outdated at the time of publication. However, we will create new versions and releases of this book to reflect on further development of components and to collaborate.

A core value of data spaces is co-creation, so this book is also a first step, but it also is open to feedback, additions and additions for subsequent releases.

After reading the book, and you feel that additions or corrections are needed, I hereby invite you to provide your feedback so that we can improve this and thus further accelerate data spaces.

I look forward to hearing from you and look forward to further development in the coming years!

Enjoy reading!

Greetings,

Gerard van der Hoeven & iSHARE team  
[gerard@ishare.eu](mailto:gerard@ishare.eu)

# Index

<b>1</b>	<b>Introduction</b>	<b>4</b>
<b>2</b>	<b>Introduction to Data Spaces</b>	<b>5</b>
<b>3</b>	<b>Layers in Data Spaces</b>	<b>10</b>
<b>4</b>	<b>Roles in Data Spaces</b>	<b>15</b>
<b>5</b>	<b>Building Blocks of Data Spaces</b>	<b>19</b>
<b>6</b>	<b>Variations and options within Building Blocks</b>	<b>31</b>
<b>7</b>	<b>How to start with Data Spaces and Building Blocks</b>	<b>46</b>
<b>8</b>	<b>Interoperability of Data Spaces</b>	<b>51</b>
<b>9</b>	<b>Data Spaces Examples</b>	<b>54</b>
<b>10</b>	<b>Get started step by step with Data Spaces</b>	<b>60</b>
	<b>Conclusion</b>	<b>67</b>

# Introduction

**Data spaces are emerging, developing and gaining more visibility in Europe. However, questions arise about how to build a data space, where to start, how to select the components, and what lessons can be learned from existing or unsuccessful projects.**

This cookbook aims to present the concept of data spaces in a simplistic manner for a broad audience, without using obscure terms or scientific jargon. It focuses on real-life applications that are relevant in today's economy.

In a few steps we will guide you through the philosophy behind data spaces, the building blocks required and their practical implementation. We will dive into the building process by providing step-by-step explanations, drawing from the experiences of players who have already gone through this journey.

The iSHARE Foundation team has authored and developed this document. The iSHARE Trust Framework, established in 2015 within the logistics sector, offers crucial building

blocks for the exchange of sensitive business data within data spaces. It was initiated due to the realization that data sharing was insufficient, despite its evident benefits. Since then, numerous parties have actively worked on practical applications of data sharing using the iSHARE Trust Framework, resulting in various degrees of success.

Based on these experiences, we have compiled this cookbook, outlining the necessary steps to design and achieve practical and meaningful data space applications. These data spaces empower data owners by shifting the control over data from central platforms back to them, allowing them to retain control over this new valuable resource.

**But what exactly is a data space?**

# Introduction to Data Spaces

A Data Space can be defined as an environment where data is shared among trusted partners who adhere to the same standards and guidelines for sharing and storing data within one or more ecosystems. This term first emerged in 2017 through a collaboration between the team of International (formerly Industrial) Data Spaces Association (IDSA) and the European Commission (EC). They were convinced that “big tech”, the major technology players from the United States, would play a disruptive role in the European economy.

An example is Amazon, which, due to its market size, can determine who makes a profit from a product and who has the right to transport those products.

Numerous platforms have attained a dominant position in various industries and markets. This has led to parties becoming dependent on these platforms, forming powerful blocks that undermine market forces between them. Consequently, data exchange programs without an intermediary platform have emerged, with key terms including:

- **No middle man:** no powerful (commercial) entities that dictate the rules within the data space.
- **Data sovereignty:** the ability of individuals or organizations to control their own data.

- **Standardized** access and exchange.
- **Open**, interoperable, non-discriminatory and federated **for all players**.

This line of thinking has led to the development of standards that facilitate secure and safe data exchange among parties. These standards establish clear agreements regarding security, certification, standardized technology components, and more.

By using the standard frameworks such as the IDSA Reference Architecture (RAM), iSHARE or Gaia-X, parties can establish exchanges based on mutual trust. It is important to note that all technically involved parties must have a strong level of IT maturity for successful implementation.

The data holder usually is not technically involved. This requirement applies to various industries, including the automotive sector. As a result, major German industrial parties have adopted the IDSA Reference Architecture as a standard.

Most data spaces were not yet open and federated in the initial phase, and they are still forming new marketplaces with a limited group of strong and influential players.

The rapid emergence of cloud providers offering software services to both SMEs and large corporations has created an opportunity for SMEs to participate in this ecosystem without requiring a large IT department consisting of hundreds of people. As a result, the European Commission and other players recognize the importance of open data ecosystems that enable SMEs to participate. That's where the need for a "trust framework" emerges where SMEs organizations retain control over their data, while certified application suppliers carry out data-related tasks according to the SME's preferences. This approach requires a shared foundation of trust, whereby data cannot be shared with third parties without permission, and all ecosystem players must reach acceptable agreements regarding the type of information being exchanged. Further details on this topic will be discussed in the upcoming chapters.

All these agreements form a system of agreements for an ecosystem with a shared objective of data exchange. These agreements, when combined with data sources and applications, establish the foundation of a data space.

Data spaces are not rigid divisions or sharp subdivisions among parties, but rather flexible collaborations between players who share the common objective of data exchange. The fundamental concept is that every organization, by definition, operates within multiple spaces or systems due to the existence of diverse interests, applications and variations.

**For example,** In an agriculture data space, a tomato grower is responsible for tracing the raw materials and pesticides used and thus crucial to the supply chain. At the same time, the grower is connected to an energy data space because they own a large combined Heat and Power Plant, which generates heat and electricity for their greenhouse, resulting in CO<sub>2</sub> emissions. Data must be supplied to both energy network operators and potential customers.

Furthermore, the grower is involved in the logistics data space, as their own trucks are used for transportation between the greenhouse and the market where products are sold.

Each of these data streams has multiple requesters and follows its own standards, forming distinct data spaces.

Any party that wishes to request data from the tomato grower can do so within their respective data space. For example, the government can use the energy data space, customers and suppliers can access the agriculture data space, and the logistics data space can be utilized to retrieve data from the grower to calculate the CO<sub>2</sub> footprint.

Currently, the grower must perform several manual actions to obtain data from various systems and share it with stakeholders. With the implementation of data spaces, these inquiries can occur digitally. Parties within a data space will be able to ask each other for information, trust each other, and respond accordingly due to clear agreements that have been established. The grower will at all times be able to grant explicit permission for the use and exchange of their data, even if it is stored in a service application instead of

their own systems. **The relevance** of data spaces lies in the ability to leverage the abundance of unused data scattered across multiple locations, transforming them into valuable sources. This is under your control as an organization.

The European Commission has proposed a generic subdivision/framework for data spaces in order to establish a general structure.

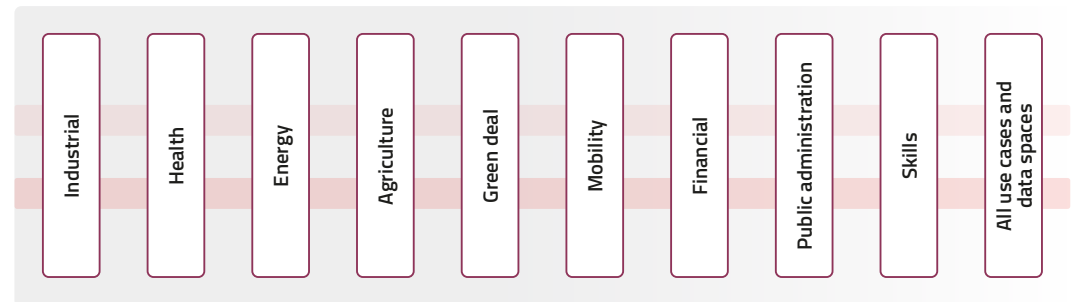


Figure 1: The data space domains according to the European Commission

This does not imply a 1:1 correspondence with parties, nor does it represent an exclusive structure. As the committee also explains, envision a bubble bath where bubbles are stacked on top of each other, and bubbles also form within.

What are the necessary steps, components, building blocks and agreements to make this practical?

## 2.1 Overarching data spaces according to the European Commission

The framework outlined by the European Commission describes nine data spaces that aim to increase the availability of data for use in the economy and society. At the same time, it ensures that the companies and individuals generating the data retain control.

### Industrial

The Industrial data space supports the competitiveness and performance of the EU industry, enhancing the potential value of using non-personal data in production.

### Health

The Health data space is essential for advancements in disease prevention, detection, treatment and informed decision-making to enhance healthcare systems.

### Energy

The Energy data space promotes greater availability and cross-sector data exchange in a customer-centric, secure and reliable manner. This facilitates innovative solutions and supports the decarbonisation of the energy system.

### Agriculture

The Agriculture data space focuses on improving the sustainability performance and competitiveness of the agricultural sector through the processing and analysis of production and other relevant data. This enables accurate and customized application of farm-level production approaches.

### Green Deal

The Green Deal data space aims to use the great potential of data to support the Green Deal priority actions related to topics such as climate change, circular economy, pollution, biodiversity and deforestation.

### Mobility

The Mobility data space aims to position Europe at the forefront of intelligent transport system development, including connected cars and other modes of transport. This data space facilitates access to pooling and sharing data from existing and future transport and mobility databases.

### Financial

The Financial data space stimulates innovation, market transparency, sustainable finance and integration, supporting the European companies and a more unified market.



## **Public administration**

The Public administration data space aims to enhance transparency and accountability in public spending, spending quality, combat corruption at both EU and national levels, address law enforcement needs, and support the effective application of EU law. It also promotes the use of innovative 'gov tech', 'reg tech' and 'legal tech' applications to assist practitioners and other public interest services.

## **Skills**

The Skills data space aims to reduce skill mismatches between education and training systems, and labor market demands.

## **Alle use cases en data spaces**

Use cases demonstrate how research applies to real-life challenges and are implemented in collaboration with industry partners worldwide.

# Layers in Data Spaces

To get a clear picture of what is required for a data space, it is important to start with a complete picture. After all, this is our ultimate objective. As stated earlier, the main goal is to share data in an open, federated and interoperable manner, for new or old uses of data without power players. A data space relies on several layers that are essential for realizing practical applications:

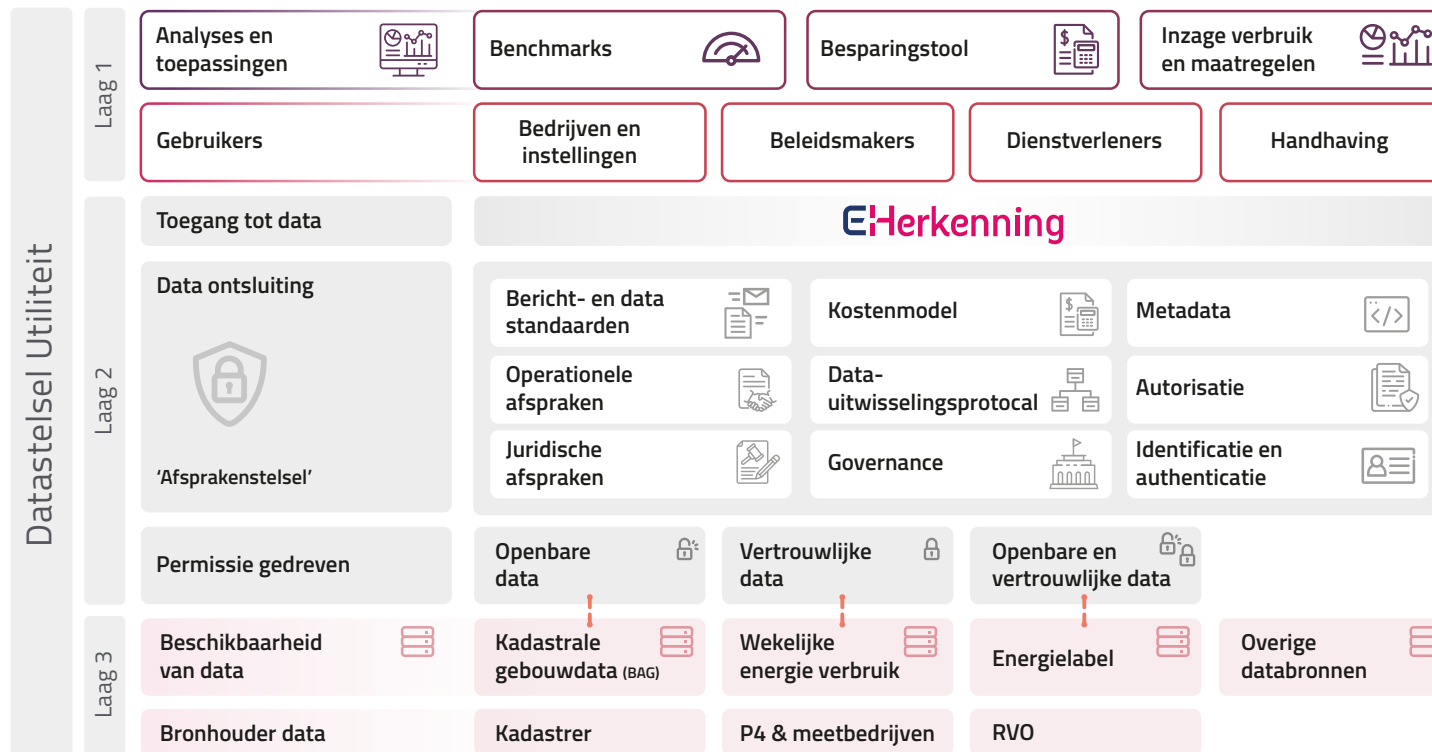


Figure 2: Data space non residential energy, RVO

### 3.1 Layer 1 – Applications and Users

It ultimately revolves around applications that use data effectively. Every successful data space, although currently limited in numbers today, begins with a real application where data exchange directly impacts the involved companies and players. This application, which is valuable to the users, also leads to actual data demands and the necessary resources for that application. This drives the pragmatic start of a first data space with real use.

This data exchange is powered by multiple service providers to ensure an open and federated ecosystem. Chapter 4 provides further information on the roles of these service providers within data spaces.

Many projects have failed due to relying solely on the hypothetical value of data without having a practical application for the information in question.

To illustrate practical applications, consider the following application examples:

#### Example Sustainability Reports

The Dutch government initiated a data space to simplify and secure access to energy, building and energy label data

of utility buildings. This initiative was part of the Climate Agreement in 2019.

Within this data space, authorized companies can exchange data by obtaining permits. This allows for direct data collection from energy network operators or metering companies. Authorized service providers can effortlessly generate reports and provide relevant advice for these companies, without the need for manual data exchange.

The required data service in this case is straightforward: What is the electricity and gas consumption of business premises “X” belonging to business “Q” on a specific day?

By combining this clear data service with publicly available information about business premises from the Land Registry and Energy Labels, powerful reports and insights can be generated.

#### Example CO<sup>2</sup> Reports and Advice

The Smart Freight Centre and the Topsector Logistiek are pioneers in reducing CO<sub>2</sub> emissions from the logistics sector. Additionally, organizations aim to accurately report their CO<sub>2</sub> emissions to government authorities. These efforts are converging in a data space being developed for CO<sub>2</sub> reports from clients to transporters (shippers).

Within this data space, the Smart Freight Centre and the Topsector Logistiek have defined the necessary data service for generating CO<sub>2</sub> reports. By establishing a standard data service in collaboration with Transport Management systems (TMS) providers (the planning tools for trucks), and AI software suppliers for CO<sub>2</sub> reports, the latter can utilize explicit exchange authorization data on behalf of their mutual consumers to provide harmonized reporting to shippers. This ensures that sensitive business data remains secure and does not end up in unintended locations.

### **Example Traffic and Water Management Inspection (IVW)**

Inspections, although inconvenient, play a vital role in ensuring road safety. The IVW monitors the trucks that operate in the Netherlands every day. Currently, random checks are performed on trucks for road inspections.

However, the implementation of a data system would significantly enhance efficiency in this process. As mentioned before, TMS platforms would allow the IVW to collect information on the content and weight of transported goods. This data can then be used to create a predictive model for identifying high-risk transports, enabling targeted inspections. By doing so, transparent “no risk”

### **Example: Pesticide Advice**

AgroTrust is a project focused on enhancing the traceability and certification of agricultural products throughout the entire production chain, from the farm to the end consumer.

It aims to establish a transparent certification solution for genuine agricultural products, strengthening trust between consumers and the entire food production chain.

By doing so, farmers and producers can guarantee food security for every citizen. Additionally, the project provides a tool for farmers to effectively manage crop protection treatment activities on their fields and track their progress. This is achieved through the utilization of Blockchain technology, enabling compliance with all legal obligations at both national and European levels.

transports could be exempted from inspections. This approach benefits all parties involved: the IVW requires fewer personnel, transporters experience reduced downtime, and more offenders are caught. Consequently, data sharers are rewarded for their participation.

## 3.2 Layer 2 – Data Space Building Blocks

Building blocks are essential for implementing the examples above, particularly for exchanging in an open and federated manner. There was originally the 9 block model developed by INNOPAY for the Ministry of Economic Affairs in the Netherlands (Minister van Economische Zaken en Klimaat). This model was further expanded into a 12-block model as part of the OpenDEI.eu program, funded by the European Commission.

This block model provides a rough indication of the necessary components and agreements required for each data space:

- **Data Interoperability:** What language do we speak to each other?
- **Data Sovereignty and Trust:** How can we ensure that “our sensitive data” only goes where we authorize it to, and not anywhere else?
- **Data Value Creation:** How can we effectively use and monetize data?
- **Data Space Governance:** How do we maintain an overview of agreements and ensure a well-structured environment?

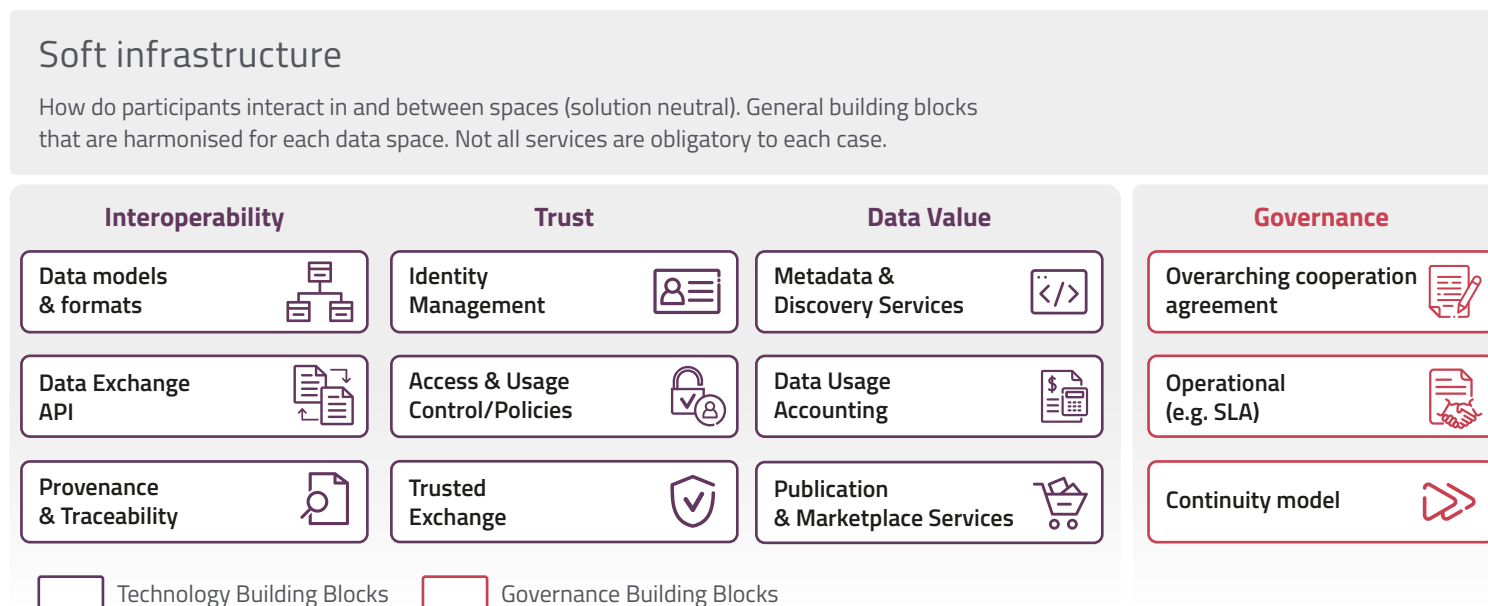


Figure 3: Data spaces building blocks, OpenDEI

These pillars are further divided into components. More detailed information about these pillars is provided in Chapter 5.

### 3.3 Layer 3 – Data Availability

Naturally, data is required for a data space. It is important to note that there are several classifications of data. Each data space has its own treatment as well:

- **Public/Open data:** Access and usage control are not necessary.
- **Licensed data:** Access and usage control are based on conditions, mainly through digital licenses via marketplaces.
- **Business confidential data:** Specific data access policies per organization are required due to confidentiality and competitive sensitivity.
- **Consumer confidential data:** Specific data access policy per organization is required for privacy reasons.

These data sources do not automatically “open fully”. Access, software and API authorization are required for all these data sources. That is why we work in data spaces with Data Services, which involve defined data service providers that deliver valuable data sets transactionally, through streaming or by other means.

# Roles in Data Spaces

In order to act in a truly open and federated manner, roles have been defined by various frameworks (such as IDSA, Gaia-X and iSHARE) to clarify the functions of everyone involved in the exchange of data. For each role, different elements (technical, legal, business, functional and operational) apply, depending on the choices of the building blocks to be used.

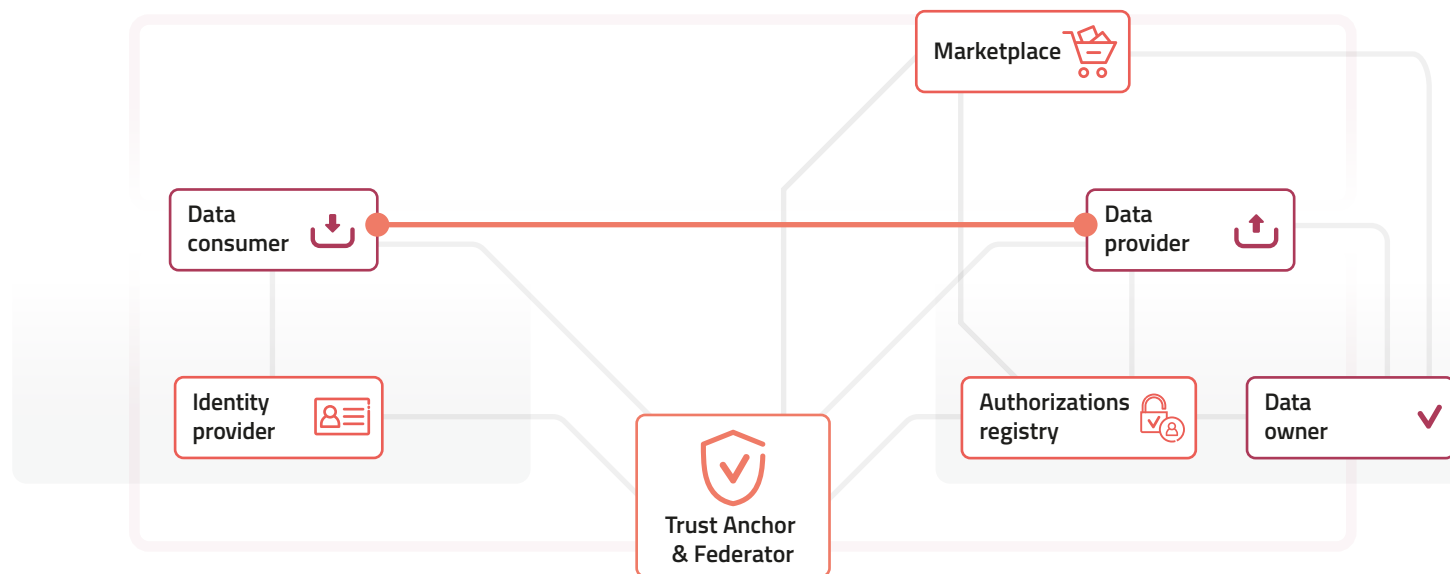


Figure 4: Roles within a data space, iSHARE

These roles fulfill specific elements of the building blocks and serve unique purposes. However, parties can also play multiple roles simultaneously in a data space.

Roles define what a party may, must and can do in the exchange of data within a data space. This information is recorded in the participant register of that data space, to ensure a continuous overview of the different players involved.

In this chapter general names are assigned for the roles, but in practice, these names and interpretations can vary for each data space.

## 4.1 Data User

An organization that requires access to existing data and receives authentication, authorization and a license to use this data is also referred to as a Data Consumer, Service Consumer or Data User. These designations roughly indicate the same meaning within the different standards.

The data user visits, retrieves, or edits the data. The user has implemented the data standard of the data space and can process it within the confidentiality agreements, signed contracts and participant registration.

Data users can be both end users and processing players who, for example, create another data product using this data, such as the Central Bureau of Statistics (CBS), which collects data to provide reports for third parties.

## 4.2 Data Supplier

When an organization provides a Data Service, such as a SaaS solution from an ERP supplier, and grants access to planning data, the SaaS ERP supplier fulfills the role of Data Provider.

Data providers offer data that can be shared with permission from the rightful claimant and have implemented an interface for data consumers to access the data. Before sharing the data, data suppliers first check with an authorization register or service (internal or external) in which the data owner has recorded permission for the data exchange.



### 4.3 Data Owner

In some cases, the individual or entity that controls the data may not have the rights to make decisions regarding that data. This is why the role of the Data Owner exists: it represents the party that possesses the rights to determine what happens to data.

This role is also referred to as “Entitled Party” or “Data Owner”, such as the owner of a vehicle, building or sensor.

The Data Owner is registered within the data space to safeguard its interests, but mainly relies on service providers (such as data users, authorization registers and data suppliers) for the exchange of data. Most data owners do not directly engage in the data exchange process themselves.

### 4.4 Identity Provider

An Identity Provider offers services for creating, maintaining, managing and validating identity information for parties involved in the agreement system.

Organizations are increasingly adhering to European standards for identification and security. Several organizations have been established for the purpose of validating organizations and issuing digital certificates.

The European movement in this area is focused on enabling organizations in data spaces based on eIDAS, euID and EORI to identify themselves and be easily discoverable. Most Identity Providers already offer these identification options.

### 4.5 Authorization Provider

Data sovereignty or data autonomy for data owners is necessary to unlock the full potential of data in platforms and longer chains successfully.

Authorizations on data are crucial to data sovereignty, under the control of the data owner.

Authorizations are stored with Authorization Service Providers, which is different from application permissions. Data authorizations/permissions control the access of third parties (other organizations) to data from you as a data owner. Application permissions grant employees access to specific parts of an application.

Authorizations can be managed by one or more authorization register players or can be established by the data owner.

## 4.6 Trust Anchor & Federator

A Trust Anchor serves as a guarantor, and can be a certifying authority within a data space. It evaluates whether parties comply with the agreements within a data space. Examples of trust anchor applications/roles include the iSHARE Satellite, IDSA's Participant Information Service, and Gaia-X's Clearing House. This role is often fulfilled by the Data Space Authority, which acts independently among the players of the data space.

Strict procedures are employed for the registration of an organization's verifiable credentials. One of these credentials is registration of the public key of an organization's certificates.

This key is used to verify digital signatures and linked to the trusted list of "certifying authorities" for validation purposes.

In addition to these digital credentials, there are other credentials, such as confirmation of validated certification of roles and validated contracts.

## 4.7 Marketplace

If accessibility to data and data offerings is a challenge, a Marketplace component becomes relevant in a data space. With the help of Marketplaces, parties can make data available under specified conditions to facilitate direct payment. An example of a Marketplace is the Fiware Marketplace in i4Trust.

# Building Blocks of Data Spaces

For the design of Layer 2 – Data Space Building Blocks, we refer to the various pillars discussed in Chapter 3.

## Soft infrastructure

How do participants interact in and between spaces (solution neutral). General building blocks that are harmonised for each data space. Not all services are obligatory to each case.

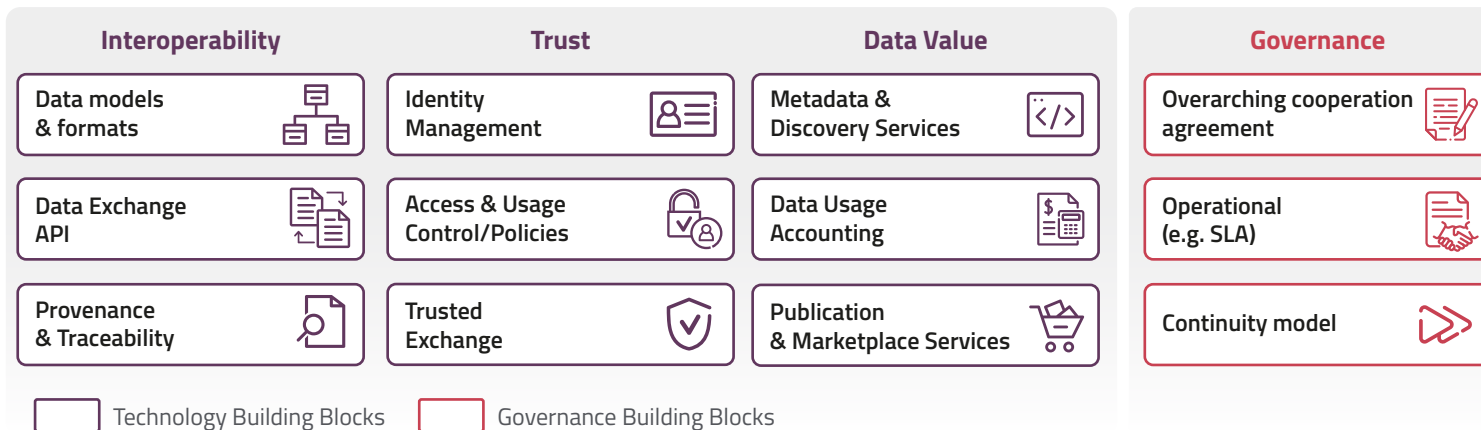


Figure 5: Data spaces building blocks, OpenDEI

## 5.1 Data Interoperability

### Data Models and Formats

This building block establishes a common format for data model specifications and data representation. In combination with the Data Exchange APIs building block, it ensures interoperability between participants.

In data exchange, the starting point is the data itself and the agreement on the languages that will be collectively used. Which data models are used, and what attributes in the data will be exchanged between parties?

This requires data models that already exist along with new ones. An important factor in designing this part is interoperability. Therefore, W3C has introduced the principle of Linked-Data (URL), where data models have a direct reference per field to a harmonized reference/meaning.

Another important factor is reducing the “whole data model”, which often consists of hundreds of fields, to services that can actually be used for data exchange. As previously discussed in the example of Sustainability Reports, the data model of the energy sector consists of 260 objects stored in the smart meter. However, for this data space, only 7 attributes are used for the first data service which is based on the same field names from the model.

### For example

In the Logistics data space, people refer to a **driver**, which denotes a **person** and corresponds to the standard for a person with a specific role. However, in the Health Data Space, this individual is referred to as a **patient**. The utilization of Linked Data based semantic models enables communication between different domains.

Subsequent services based on the same model can be implemented in the same data space as the next step for parties. We will discuss this further in the chapter on **Variations and Options Within Building Blocks**.

Unfortunately, these models are often reinvented in different countries for the same segments, which is understandable due to a lack of collaboration across different domains. With the emergence of data spaces spanning various playing fields, the need for standardization has become even more crucial. Moreover it offers direct savings for all involved players by reducing the number of required integrations. Initiatives such as SmartDataModels.org and Schema.org are driving the development and maintenance of standardized data schematics for different domains.

## Data exchange API

### What standards are necessary for data exchange?

This building block facilitates the sharing and exchange of data (i.e., data provision and data consumption/use) among participants within the data space.

The data services are logically situated with the data service providers. The Exchange API is the actual technology that enables the delivery of a data service, based on a standard chosen within the data space.

Each Exchange API can be queried for a specific defined service, targeting a particular organization with a specific query purpose. The definition of the API of which service is offered follows from the metadata for the various services.

The technical standard is crucial, including the use of SOAP, REST, EDI, as well as FTP and email, which are still utilized as data exchange models. However, some of these models are better suited for data spaces than others. Active authorization for data access is a requirement for data spaces to function effectively. This, for example, is not feasible with email and EDI.

It is important to note that not all data is available online or directly connected to the data space. Recently, data proxies or platforms have started to emerge. These proxies don't store data themselves but act as intermediaries between less digitally-oriented sources and the data space. **DEFLog** in the logistics sector is an example of such a platform. These proxies ensure that the data owner maintains control while bridging the gap between traditional data sources and the data space.

Another example is the iSHARE data hubs (platforms such as Portbase) where data is stored and processed, making data sharing more streamlined.

### Provenance & Traceability

This building block provides the means for tracing and tracking throughout the process of data provision and data consumption. It forms the foundation for several important functions, ranging from identifying the lineage of data to maintaining audit-proof logs of transactions. It also enables the implementation of various application-level tracking use cases, such as tracking product or material flows within a supply chain.

In addition to the components for data access and user supervision, there are cases where additional monitoring of authorizations and datasets is necessary. While technical access control and usage control already address this, this building block adds further capabilities.

The provenance component focuses on demonstrating the origins of data. It involves metadata associated with records that describe the data's source, any changes made to it, and details that support the trustworthiness and validity of the data. Data provenance is crucial for detecting errors within the data and attributing them to sources.

In other words, data provenance helps answer questions such as "why was the data produced", "how was the data produced", "where was the data produced", "when was the data produced", and "who produced the data".

The traceability component serves as a monitoring and control element. This information is recorded, for example, in the Authorization register of parties, documenting who has used which authorization and at what time. It enables the detection of system overloads, authorization errors, or attempts to access data without proper authorizations.

Additionally, it monitors the flow of transactions throughout the entire chain to ensure compliance with the rules of the data space.

In addition to these aspects, data spaces can be envisioned where all transactions are monitored and settled, or where tracers are set for longer transactions. This could be implemented using blockchain technology, another token structure or verifiable credentials. Such advanced features offer possibilities for more sophisticated data spaces.

## 5.2 Data Sovereignty & Trust

### Identity Management

The Identity Management building block enables the identification, authentication and authorization of stakeholders working within a data space. It ensures that organizations, individuals, machines and other actors are provided with recognized identities that can be authenticated and verified. This includes additional information provision, which is used for authorization mechanisms to enable access and usage control.

Entities in data spaces that need to be identifiable could be anything, and such entities are identified with a number. However, there are few nuances that are important to recognize and acknowledge:

- A natural person may have several numbers, such as a personal number from the government, like their BSN.
- A legal entity has various numbers such as a Chamber of Commerce number and a VAT number.
- Employees of a company and authorized representatives (directors, beneficial owners, etc.) are a combination of the above, having the authority to sign or approve actions on behalf of an organization, and at different levels.
- Asset Identity, an important factor within data spaces, refers to the digital twins of physical objects, such as business buildings, vehicles and containers.

In all these cases, tight coordination is essential so that identification can be effectively used. One of the core elements for interoperability between multiple data spaces is the use of unified identities across these data spaces. Regardless, organizations often operate in multiple segments and therefore, in multiple data spaces. Finding parties in an unambiguous

way is an important decision as it is crucial for both trust and findability. We will revisit this topic in chapter **Get started step by step with Data Spaces.**

The different Levels of Assurance for verifying companies' identities are a choice within the data space. An example of assurance levels for identification is E-Recognition, where higher levels require greater assurance and testing.

### **Human Identity Provider**

Initially, organizations process the data of individuals, who have their own unique identities. To access this information and verify their identity, Human Identity Providers are required. One example of such a provider is iDIN.

### **Organisation Identity Provider**

An important part of the transactions will be on behalf of organizations, whether or not performed by an individual acting on behalf of an organization. Organizational Identity providers exist to fulfill this role.

Organizations are increasingly complying with European standards for identification and security. Many organizations have already been established for this purpose to validate an organization and issue a digital certificate.

The European movement in this area aims to identify and locate organizations in data spaces based on eIDAS and EORI. Most organizations already possess one or more electronic means of identification, used for submitting tax assessments, applying for subsidies, etc. E-Recognition is an identity tool that provides certainty regarding the identity and authorization of an organization.

With the EORI number, economic operators are identified in the same way by customs authorities. This provides efficient benefits for both market participants and custom authorities. These EORI numbers are validated based on a validation request issued by the European Commission.

### **Trusted Exchange**

This building block facilitates reliable data exchange between participants, assuring participants within a data exchange transaction that other participants are who they claim to be and that they adhere to defined rules/agreements. This assurance can be achieved by applying organizational measures (e.g., certification or verified credentials) or technical measures (e.g., remote attestation).

The trust anchor is the foundation of the data space, where the most valuable components of digital validation come together. In this context, a trust anchor is an authoritative entity for which trust is assumed and not inferred.

A digital registry records the role played by organizations, along with their certifications and legal coverage in the data domain. Such a registry is a natural tool to be used by a trust anchor. The credentials of parties can be digitally verified, forming an important building block and startingpoint in any data space.

Furthermore, the trust anchor addresses fundamental discoverability issues, such as determining the Authorization registry where parties have permissions and the location of the party's data services. This will be revisited in the "Discovery" section but is part of the same concept.

### **Access & Usage Control**

This building block ensures the enforcement of data access and usage policies defined as part of the terms established when data sources or services are published or negotiated between providers and consumers. Typically, a data provider implements data access control to prevent resource misuse, while data usage controls are implemented on the data user's side. In complex data value chains, both mechanisms are combined by prosumers (a combination of producers and consumers). Access control and usage control are based on identification and authentication.



Data autonomy of data owners is the core principle of data spaces, and it encompasses two basic principles:

- Access control, determining who is allowed to read which data service under what conditions.
- Usage control, specifying the purpose for which this data may be used.

An Authorization registration is a resource where the data owner can register these two basic principles. Registration can be completed in an independent register, known as the Authorization Registry. However, registrations can also be stored and managed at other locations, depending on the data space. Therefore, authorizations, permissions, or policies are stored with authorization service providers.

It is crucial to understand the concept of these permissions, as they determine which party can access specific parts of third-party data and the reason for it.

Specific configurations for policies can be established in data spaces. However, the choice of how to implement these configurations lies with the companies. For example, they can use open-source components developed in-house or rely on one or more application service providers.

For each data service, it first must be checked whether a number of policies are recorded in the authorization process (in a data service definition) and whether it is possible to make them retrievable via the Authorization Registry.

### Example

**Company X** authorizes **Statistics Netherlands** to request the data service “**employees**” from data supplier **Exact**, with the license to create **anonymous sector reports**, for the period **1-1-2022 to 1-1-2023**.

These authorizations are created based on questions. In the example above CBS asks the Exact service whether they can access the number of employees of all their customers. Subsequently, all these companies receive a notification in their own Authorization Register, asking whether they want to authorize Exact to share the data with Statistics Netherlands. This process can prevent the need for sending 100,000 letters per year.

This approach makes it extremely easy for data owners to monitor permissions, and it is straightforward for many parties to query this data while ensuring strict monitoring.

## 5.3 Data Value Creation

Once data is present and access permissions have been established between trustworthy parties in a shared data space, the next step is to figure out how to access it.

The following components and interpretations are required for this:

### Metadata & Services Discovery

This building block includes publishing and discovery mechanisms for data sources and services using common descriptions of sources, services, and participants. Such descriptions can either be domain-agnostic or domain-specific.

Metadata in this context describes the definitions and quality elements (claims) of this data service. This can differ significantly per data space, but it defines, for example, which standard is used by a service and which certainties are part of the quality of the data service.

In addition to the “discovery of the services”, another search element is required that indicates which parties are registered in data space x or y. iSHARE offers the “parties” data service for querying active players in a specific

data space. This enables every player in the data space to access the information.

Firstly, the findability of the services is important. This requires a starting point where you begin your search.

Thereafter, it is important to know what properties and qualities this data service has, for example, the standards and variables used. Within iSHARE this is called the Capability, which gives the description of the data service, including the data standard, service location, server security aspects, etc. In Gaia-X, this is called “service description”.

Some data space models use a search function that queries each data service gradually, looking for a specific service. It is an option, but requires a relatively large amount of time and effort although it is useful when a marketplace or satellite is unavailable in the data space.

In data spaces with one or more trust anchors/data space authorities, this is generally the starting point to find data from a specific organization. In iSHARE, this is called the “capabilities endpoint”, which refers to a data service that provides an overview of all active data services of a party, with a metadata set per data service.

## Publication & Marketplaces

To support the provision of data sources and services under defined conditions (for example, price or nationality), marketplaces can be set up in a data space. This building block supports the publication of these offerings, the management of processes related to the creation and monitoring of smart contracts (which clearly define the rights and obligations for the use of data and services), and access to data and services.

If the data space lacks a basic “value agreement” determining payment responsibilities, for instance, if it doesn’t provide savings for all or a collective benefit, the Marketplace platform can be a suitable solution. With the help of marketplaces, parties can make data available against financial conditions and pay directly. The FIWARE marketplace in i4Trust is an example of this.

On the data marketplace, data services including their metadata, are made available under specific **conditions**. These metadata fields contain the definitions and quality elements (claims) of this data service. These conditions include, for example, the price, the permitted use, and any restrictions. These conditions are generic for all parties, not player-specific. A transaction on the marketplace eventually generates a policy in “Usage and Access Control”.

## Data Usage Accounting

This building block forms the basis for accounting access to and/or use of data by different users. This, in turn, supports key settlement, payment, and billing functions (including data sharing transactions without going through data markets).

Anytime data services are used, this is registered, for example, in the Authorization Register or at other locations where the usage is tracked (for example, a transaction ledger). The usage reports on this registry can be shared as validation of the marketplace transactions. At the same time, the marketplace is autonomous in the settlement of data services, to ensure that the authorizations remain active. The marketplaces act as an intermediary in this process, but the execution of the access remains in the Authorization Registry of the organization itself, to avoid forming a new power block by the marketplaces.

### Example of Data Conditions:

Company X offers a data service that provides insight into the number of cars driving past a specific location. Data from this sensor can be accessed if the following conditions are met:

- EUR 7.50 per month
- European parties only
- Only parties registered within the Mobility Data Space.

## 5.4 Data Space Governance

We will now discuss the most underexposed theme of data spaces. Technology is widely discussed, however, without governance, a data space cannot exist.

Governance sounds like a vague term but what does it mean? Governance involves:

- Defining, setting up, and managing the data space
- Facilitating the further development of the data space
- Facilitating operational processes, such as onboarding and offboarding
- Data space monitoring (such as compliance monitoring and dispute resolution)

Governance of data spaces is generally performed by parties that have no commercial interest in the data exText. It may be the case that parties “set up” an independent body within their data space, for example, a foundation. One of the key roles for the party organizing/managing the governance is to validate the “trustworthiness” of parties in the data space. This is tested with the trust framework used by the relevant data space. A Trust Framework is a common set of best-practice rules that ensure minimum requirements for security, privacy, identification management, and interoperability are met through accreditation and governance.

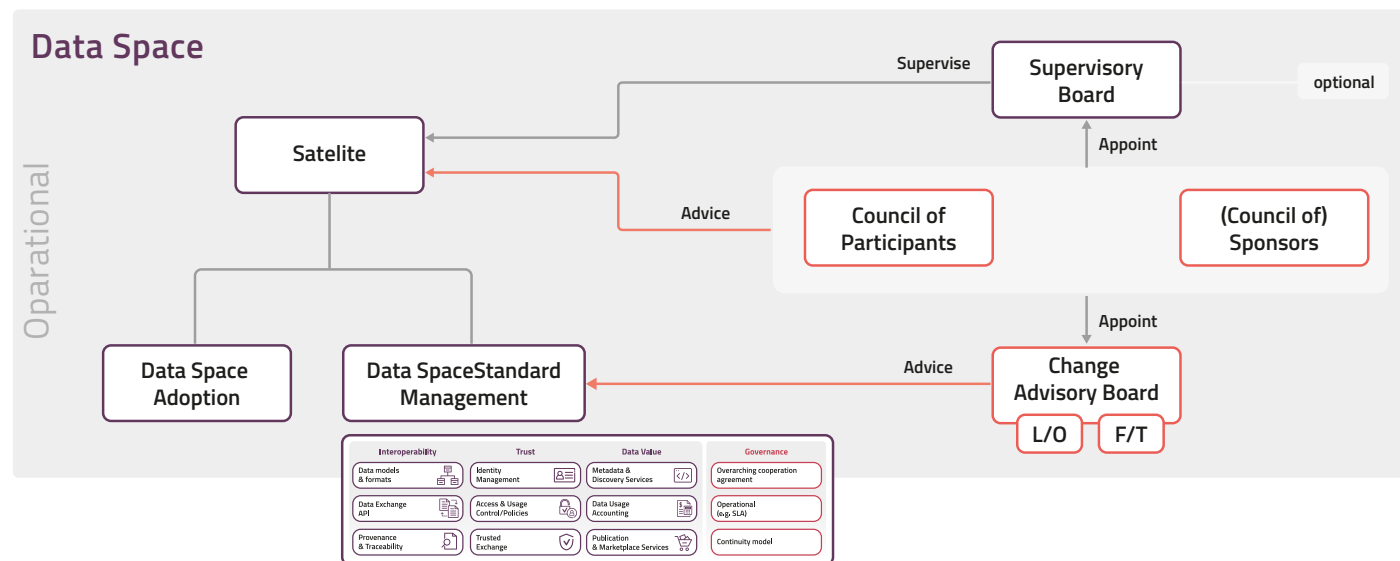


Figure 6: Organizational chart  
iSHARE Foundation, iSHARE

This party adds other parties to the ecosystem, but also organizes the “co-creation” process of forming a data space. The latter may sound a bit unclear, but as indicated in chapter 3, we want to prevent a central party from defining the conditions of the data space. By co-creating in the development and management of the data space, a level-playing field is created among all market parties who jointly influence the legal condition of the data space.

Individuals within the shared data space are able to contribute to its governance and have a voice in determining its features and capabilities.

Through this approach, the participants of the data space have direct influence on all definitions and agreements within it.

### **Overarching Cooperation Agreement**

All data space participants must agree on functional, technical, operational and legal aspects that are relevant to the data space. While some agreements are reusable in a generic or industry-specific way (rulebooks, for example), others are use-case specific.

By using a Legal “umbrella agreement”, from a Trust Framework such as iSHARE, for a number of aspects of data spaces, legal interoperability is created between all data spaces that use the same framework. This enables the use of each other’s data sources with shared certainty regarding user rights, liability and protection.

#### **For example**

DVU reports on business energy usage, while DSGO has applications for making buildings more sustainable, which require valuable access to energy information.

Both parties use the same trust framework, enabling them to find and query each other on an organizational level (e.g., iSHARE Satellite) and establish trust on a legal level.

With this method, DSGO can send a request to a data service from DVU, and via the existing authorization register service of the data owner (assuming there are no legal restrictions), DSGO can request data from this data owner at other services without the need for reconfiguration.

A practical and proven basis for this is the iSHARE Trust Framework and the legal system.

The signature of this legal system is registered in the participant register, making it possible to verify which party has specific coverage for its specific data space.

## Operational

If there is a complete basis of trust between parties, the question arises of whether the services are operationally mature. Even if everything is arranged regarding data models, trust and authorization, if the service is unavailable 50% of the time, participants of the data space cannot rely on it. For example, in a data space where collection times of goods are exchanged, it is crucial for the driver to know with certainty what time they are expected somewhere. Otherwise, the data space is of little use in everyday operations.

Therefore, there is a need for clear and committed Service Level Agreements (SLAs) among the players in the data space. iSHARE provides a basic SLA as part of the onboarding process in the framework. However, each shared data space is unique, allowing for flexible growth and expansion.

However, when it comes to the domain of mobility traffic flows, relying solely on a 99.7% SLA is insufficient.

## For Example

In the realm of business energy data, no additional Service Level Agreement (SLA) has been implemented, and the existing SLA stands at 99.7%. While occasional unavailability of a service is possible, stability can be achieved by persistently resending requests.

Even if everything operates flawlessly, there remains a 0.03% chance, equivalent to one entire day at a traffic light, where one may experience a service delay during 365 days.

In addition to SLAs, it is important to make good agreements about access, incident management, release management, and reporting. The procedures described in the iSHARE Trust Framework can serve as a basis for this. Refer to this [link](#) for further explanation of iSHARE operational processes.

## Continuity model

Naturally, the data space must be viable and sustainable. That is why certain choices are necessary within the data space, such as decisions regarding the financing of the data space itself. This includes determining who pays what, when and to whom. Membership fees, subscriptions to data services, and other models may be considered. More examples will be shared later in the book.

# Variations and options within **Building Blocks**

Now that we understand the meaning of the different blocks, the next step is to choose the right building blocks that are essential for the data space. When, how, and to what extent do we use them?

## 6.1 Data Interoperability

Data models and formats are crucial for federated collaboration, as explained earlier. However, the starting point for achieving this should not be the data model itself. Instead, it should be the application that requires access to federated data. This approach ensures that the specification is guided by the application's needs.

Keep in mind that before opening up any data field within a shared data space, development work must be done. If you want to unlock a data model with 500 attributes, it will require a significant level of development. However, it may be sufficient to use only 5 fields for the application. In other words, start small and scale smart.

The sequence should be as follows:

- Determine the data requirements
- Map these requirements to existing standards
- Choose an appropriate standard
- Ensure it is mapped to Linked Data

To illustrate, here is an example:

### Example

The data space NL energy for companies begins by opening up 7 fields, for which descriptions and required field names have already been provided by the application. Instead of diving into the definition phase immediately, the team first examines SmartDataModels.org and Schema.org to check whether there are already existing linked data-based standards.

Fortunately, such standards do exist, and as a result, interoperability with the Spanish data space in the same domain is immediately achieved.

## **Data Exchange API (Application Programming Interface)**

Unlocking data always requires development, as many applications face integration and software development costs. That is why the starting point of a data space API set is: “Where is the data of most companies for this starting service?” and “Which standard interfaces are already in use?”.

In addition, harmonization between software parties is an important factor. In the foundation of a data space, it is essential to ensure that majority players from the same segment provide APIs in the same way.

This criteria of access is often a threshold that must be taken into account. This is not due to the complexity of preparing the “standard API” that makes the data available with the same standard, but so that this service can be used by any party, as long as there is authorization.

Software parties have benefited significantly from custom integrations that are no longer necessary with the help of data space principles. Therefore, this presents an element of market failure which requires many conversations or funding. In addition to the evangelization efforts, it is reasonable for software suppliers to seek compensation for enabling data space connectors for various parties.

The choice of the technology of the API standard, therefore, follows from the above, whereby the use of REST APIs is obvious.

In the last 10 years, Representational State Transfer (REST) has become a defining principle for realizing APIs. So-called “REST APIs” function for applications as websites do for people. Websites present information to people, REST APIs make applications and data available over the Internet to other applications. Therefore, the technology behind websites and REST APIs has a lot in common.

The Dutch government uses REST APIs for links with other governments, companies and indirectly with citizens, for example, via mobile apps and web apps offered by companies or governments themselves. Developers can query these REST APIs from common programming languages and frameworks such as Python, Java, Microsoft C#, and PHP.

The standard REST API Design Rules aim to bring more uniformity to how the government offers REST APIs. These rules describe the basic principles for structuring and documenting REST APIs.

The REST API Design Rules must be applied where the government uses REST APIs but do not mandate the use of REST APIs when accessing data or functionality.



In addition to the REST API, there are API Standards such as GraphQL and EDI that can be used. The choice of API standards varies in different data ecosystems. For example, EDI (Electronic Data Interchange) offers many degrees of freedom, leading to the development of various interface variants by different parties.

### Provenance & traceability

First, it is crucial to determine the necessity and relevance of this role within the specific data domain. The primary objective of this role is to ensure the proper processing of transactions, thereby establishing trust. In the digital data landscape, numerous stakeholders express concerns regarding the visibility of their data and activities to others. Hence, it becomes imperative to exercise due diligence to protect and uphold confidence, as exemplified by the logistics sector, among others.

Evidence is required for the settlement of transactions. This also applies to the example above. After all, how does the client know that a transaction has been properly executed without proof being provided?

1. It is essential to assess the data space and identify the level of **provenance required** for its applications. Take the logistics industry, for instance. When utilizing the Electronic Consignment Note (E-CMR), there are several options available for establishing provenance. One such option is central provenance and traceability, where one or more independent players within the data space are designated to handle this responsibility. In the case of E-CMR, service providers are responsible for facilitating updates and additions for the participating parties.
2. Another option is **provenance per party**, whereby each company uses its own (decentralized) E-CMR service via its own ICT infrastructure.
3. Lastly, **transaction provenance** can be embedded in the transaction itself, adding more proof to the step-by-step in the token of the transaction for settlement, without central services (verifiable credentials and presentations).

## 6.2 Data Sovereignty & Trust

### Identity Management

The degree of certainty about someone's identity is crucial in any ecosystem that works with confidential information, as also explained in Chapter 5.2.

When making decisions pertaining to identities for companies, employees and natural persons, it largely depends on the confidentiality of the data in the applications within the data space.

In iSHARE, this is defined as "Level of Assurance", indicating the level of identity reliability. This level of reliability is distinct for different applications.

The required level of assurance determines the choice of identity provider and which identities are accepted within the data space. In data spaces, we ideally reuse identities that parties have previously registered with one of the identity providers affiliated with the system, or we allow relevant identity providers that are not yet affiliated to seek connections.

### Explanation High Level of Assurance

In a data space dedicated to the exchange of health information among hospitals, a robust level of certainty is imperative. The validation process for identities and certificates during onboarding necessitates a rigorous level of scrutiny and verification.

### Explanation Low Level of Assurance

In a data space where the data service relies on building information sourced from the Land Registry, a moderate level of assurance suffices. While there is a need for transactions, the level of certainty required beyond that is relatively limited.

Existing identity providers have applications and an operational process for authenticating and validating users. Some examples of existing identity providers and components that can play a role are:

### **eIDAS Qualified Identity Providers**

eIDAS (Electronic Identification, Authentication and Trust Services) is an EU regulatory framework that regulates both electronic identification and general trust services related to electronic transactions. Launched in 2014, eIDAS is part of the European Commission's strategy to stimulate and accelerate digital innovation in the region.

More specifically, the EC calls for the introduction of an European digital identity system that gives every citizen and company a unique and verifiable reference. These can then be saved, accessed and used for multiple interactions, both online and within the EU.

Identity providers that supply qualified eIDAS identities meet the high requirements set by the EC for identities and thus provide basic security.

eIDAS 2.0, an updated version of its predecessor, has been developed to remove a number of barriers with the original framework.

eIDAS 2.0 offers the possibility to be used in both public and private domains, whereas eIDAS 1.0 was limited to the public domain. This makes a scalable identification and authentication challenge much easier in the future.

### **E-Recognition (eHerkenning)**

E-recognition is an identity framework that was originally started by government services, allowing organizations and companies to make and keep their online services secure. It was founded in 2009 and is the business variant of DigiD.

Prior to 2009, there was a lack of a secure and trust-worthy registration system for companies to engage in business with the government. E-Recognition ensures the authentication and authorization of a person who wants to purchase an online service.

E-Recognition is a key component of the comprehensive digital infrastructure that constitutes the foundation for electronic government services. E-Recognition falls under the electronic access services agreement system (ETD system), which is part of the eID system. E-Recognition complies with the eIDAS requirements and serves as a recognized European login tool.

### **Non EIDAS Identity Providers**

Secure Logistics specializes in developing secure and privacy-proof identity and access management solutions for various industries, including logistics, construction, and industry.

Their innovative products, such as XS-IDs and XS-Keys, enable the secure sharing of personal data within the work environment, giving full control over the data they share and with whom. By accepting XS-IDs and XS-Keys, businesses can establish an efficient, reliable, and privacy-proof access process.

Secure Logistics follows rigorous procedures to validate the authenticity of organizations and ensure that only authorized individuals have signed off on the access.

### **Open-source Identification Resources**

Keycloak is an open source software product that enables single sign-on using Identity and Access Management (IdAM) for modern applications and services. Users authenticate with Keycloak instead of individual applications, eliminating the need for company applications to handle login forms, user authentication, and the storage of user information. Once users are logged in via Keycloak, they do not need to log in again to access another application. Additionally, this applies to the logout process, simplifying it for users as they only need to log out once to be logged out of all platforms.

Keyrock is a FIWARE component responsible for Identity Management. By using Keyrock, a company can add OAuth2 based authorization security to its services and

applications. OAuth2 is a protocol that allows users to allow third-party websites or applications to access protected resources without necessarily revealing their long-term credentials or even their identity. The level of security provided by an Open Source component is yet to be determined. However, it is crucial that the controls and procedures for issuing Identities align with the requirements of the Trust Framework, such as iSHARE, within the Data Space.

### **Trusted Exchange**

The trusted exchange, as defined in chapter 5.2.2, is a core element within a data space, determining the interoperability of the data space with other data spaces in terms of trust and organizational details.

Depending on the scale of a data space, there are several variants for setting it up.

### **Small-scale data space with well-known players**

During the experimental phase, establishing a trust anchor may be relatively easy compared to a larger setup. However, it is still important to be careful in choosing the appropriate design approach.

Frequently, experiments and technology initiatives are initiated before the development of the data space architecture. However, this approach has proven to be

significantly inefficient as it creates conflicts between platform development and the establishment of a coherent data space. In platform development, all components, such as APIs, identification, authentication, authorization, trust anchor, participant registrations and much more, can be freely configured and are therefore often developed from scratch. Standard building blocks and interfaces are used within the domain of data spaces.

The work is effectively identical in both cases. But when converting from platform to data space, a major extra step needs to be taken.

The Trust Anchor serves as a crucial component that can prevent unnecessary duplication of work when establishing a new data space. It provides a central registration of organizations, allowing for the verification of whether parties have been reliably tested against the requirements of the data space and the chosen trust framework. This significantly streamlines the process and ensures adherence to the necessary trust and validation protocols.

For the experimental phase, registration in a Trust Anchor such as the iSHARE Satellite is a direct simplification and form of saving for the development part and a connection towards the next phase.

### **Closed data space with known players**

A data silo refers to a data repository that is exclusively controlled by a particular department or business unit, effectively isolated from the broader organization. The data within a silo is usually stored in a standalone system and is often not compatible with other data sets.

A closed data space, by its nature, does not engage in cooperation with other data spaces. In such cases, the question arises as to whether it can truly be considered a data space. However, there are several possibilities to transform this closed system into a more open one:

### **Dedicated software components**

Dedicated software means that components within the software system are custom-made to meet a customer's needs. This provides customers with the ability to maintain control over access to their information, allowing them to determine who can have access to it.

### **Reusing open-source components (iSHARE & FIWARE)**

Open-source software components consist of software that has been licensed for distribution and usage in other applications. FIWARE, Gaia-X, IDSA and iSHARE are examples that grant these licenses.

### Link with existing Identity Providers

Existing identity providers have applications and a running process for authenticating and validating users. Numerous examples of existing identity providers and components that can fulfill crucial roles have been previously cited and elaborated upon.

### Open and Federated data space

In order to operate in a genuinely open and federated manner, specific roles have been defined within frameworks such as IDSA, Gaia-X, and iSHARE. These frameworks outline the roles and responsibilities involved in the exchange of data, ensuring clear guidelines for all parties involved.

Once the relevant building blocks for the data space have been identified, the technical, legal, and operational aspects of each role must be determined. Within the data space, different parties can assume specific roles, and it is also feasible for a single party to undertake multiple roles. These roles establish the permissions and obligations of each participant in a data exchange and are documented in the participant register of the data space, ensuring a comprehensive overview of the various stakeholders involved.

The iSHARE satellite serves as the central coordinator and governance core within the iSHARE Trust Network, playing a pivotal role in a data space. The management of the

iSHARE satellite is assigned to coordinating organizations that have been onboarded by the operating participants, known as the Data Space Authority. These organizations take on the crucial role of supervising and facilitating the operations within the data space.

The organizations that function as satellites are responsible for vetting the participants within the data space and enrolling them in the iSHARE Distributed Ledger. This vetting process ensures that the party in question is a legitimate entity and fulfills the criteria established by both iSHARE and the data space, encompassing legal, technical, and other relevant requirements set forth for participating parties.

The iSHARE distributed ledger can be accessed according to the standard to identify and authenticate participants within a data space.

By registering all participating parties within the iSHARE distributed ledger or interoperable registers, a crucial assurance is achieved. This allows a party, once it has successfully passed the vetting process within its satellite, to become a participant in other data spaces, provided it meets the specific requirements (legal, technical) set by each individual data space. This verification process is performed by the satellite of the respective data space.

While utilizing the iSHARE satellite and ledger is one viable approach, data spaces also have the flexibility to adhere to alternative trust schemes that align with the same principles. The choice to utilize the iSHARE satellite and ledger is optional, as data spaces have the freedom to adopt equivalent trust frameworks.

## Access & Usage Control

Depending on the type of data space, choices can be made regarding the level at which access control and usage control are set up. The optimal choice for a data space depends on the classification of the data within that particular space.

The level of confidentiality required and the applicable legal frameworks for the data are influential factors in determining the most suitable approach. When exchanging personal data, compliance with the General Data Protection Regulation (GDPR) is crucial, while for company data, adherence to the guidelines and regulations outlined in the Data Governance Act would be pertinent (<https://digital-strategy.ec.europa.eu/en/policies/data-governance-act>).

The initial decision to be made is to establish the level at which parties will have access control to data. This access can be authorized based on the specific roles that parties fulfill. For instance, Customs, by virtue of its role, may

have access to comprehensive details related to goods. However, regardless of the specific scenario, data policies serve as the fundamental principle, enabling access to be determined for each dataset and data service.

The second decision revolves around determining the level of usage control to be implemented within the data space. Several questions arise in this regard:

1. Usage control refers to the actions an organization is permitted to take with the data once it has been acquired. This is typically governed by licenses that outline specific permissions granted to the data consumer. For instance, authorization may determine whether the data can be shared with others.
2. Considerations include the level of trust within the data space and the relevance of the data to the receiving party. These factors influence the decision on how the data can be effectively utilized and whether it aligns with the needs and objectives of the recipient.
3. The technical feasibility of the chosen approach is another crucial aspect. It is essential to assess whether the chosen level of usage control can be practically implemented within the data space, taking into account technological capabilities and limitations.

### **Explanation iSHARE Licenties**

Licenses are a crucial role in iSHARE as they enable participants to define clear permissions and restrictions. As all iSHARE participants are bound by the same contract (Terms of Use) and underlying scheme rules, they can trust that others will comply with the provided licenses.

### **Explanation of IDSA Connector with Usage Control:**

The IDSA Connector provides a dedicated space for applications that can process received data within the secure environment of the connector. This arrangement enables processing of data without exposing the raw data to the external environment. Only the processed results are shared.

Both the connector and the apps operating within it must undergo certification.

We identify three main options:

The first option of usage control is legal, where for example, the iSHARE Trust Framework provides coverage for the use of data through the use of data licenses that are anchored in the framework. This will be a comprehensive approach for a large part of data spaces because data does have to go to the user to be combined with other sources or processed as a transaction.

If participants require a specific license, they can submit an application to request its inclusion.

The second option involves implementing technical usage control, as exemplified by the use of IDS (International Data Spaces) connectors. In this approach, distinct applications operate within the data consumer's connector, allowing granular control over data usage at a technical level.

These applications do not run in the public domain of the data consumer, but instead serve as bridges between the consumer's domain and the incoming data. These applications have been checked and certified against the IDSA framework, providing assurance that technically the data cannot be sent anywhere else than indicated in the licenses.



### Example of MPC in Police Investigation

Multi-Party Computation is a production technology used to generate collaborative data insights without revealing the raw data. Cryptographic techniques, such as MPC, accelerate data access and enable multiple parties to analyze data together and draw conclusions without exposing each other's data. Calculations and analyses are performed on encrypted data, and only the final results are decrypted. With MPC, no data is transparently disclosed; only the conclusions derived from that data are shared.

Moreover, in the Netherlands, data is often stored in a closed format, inaccessible to others due to legal reasons or concerns about sensitive data being disclosed by unauthorized parties. MPC can break through these barriers and open up new investigative possibilities for the police.

The third option involves a growing area where monitoring is not conducted by the data consumer. Instead, the data provider operates a Node/Blockchain component that performs analysis directly at the data provider's location, in proximity to the dataset. In this approach, the data itself never leaves the data provider's environment, and only the outcome of the calculation is shared. This technique is known as Multi-Party Computation (MPC).

An example of this is the application of Roseman Labs or the Nodes as defined in the FEDeRATED program.

### Deployment Options

The same consideration applies in this case as well: whether to reinvent or reuse existing solutions.

An example of reusing existing solutions is the availability of ready-made open-source components for authorization. These components are already in the pilot stage, making it easy for software teams to begin utilizing them. Pilot accounts are also available with iSHARE authorization registry suppliers such as Visma, Poort8, DXC, Portbase.

By opting for an out-of-the-box version provided by a software supplier, it becomes possible to swiftly establish a pilot environment with limited resources. This approach

enables streamlined setup processes. However, it is important to note that there is also the option to develop a custom solution in-house, depending on specific requirements and available expertise.

## 6.3 Data Value Creation

The financial value creation will differ per data space. In some data spaces, data will mainly be traded, while in others, no payment has been agreed to at all.

Discovery of data offerings from players remains as the crucial element.

### Metadata & Services Discovery

As also explained in chapter 5.3.1, data spaces always have a starting service or set of services. The metadata description is essential for a good understanding of the data in the data space and the approach on all sides.

Starting with a machine-readable format isn't necessary. Many data spaces begin with a physical description of the service. Chapter 9 examines the various data spaces in operation in the Netherlands and elsewhere. The best option depends on your level of maturity, where pragmatism is helpful.

### Central Wiki with service descriptions and metadata and central service provider

As a starting data space, you begin with an ecosystem that brings together a first group of parties. A good starting point is to begin with a specification set that parties can implement in their systems and processes. This allows parties to understand each other and start with the first services.

Starting with a "paper" agreement system is a good starting point, has a low threshold and ensures good coordination. This paper structure also contains the descriptions of the services, properties of services, etc., so that developers can start quickly.

### Participant-driven Information System

A participant in an information system is often seen as a direct user, as this is an integral part of the functionality and a trust basis for the data space of the system. A participant is considered part of the information system and helps in the execution of the information processes.

The participants in this service within the data space are monitored in the same way as others, based on, for example, the IDSA or iSHARE certification procedures.

The interfaces to this ParIS (Participant Information Service) do meet the standards to ensure that parties can be found in an unambiguous manner.

### **Distributed Nodes**

Distributed Ledger Technology, like iSHARE, leverages its participant register to enable nodes participating in the network to securely store participant data in a shared database. This database can be accessed by authorized data space authorities, subject to permissions granted by participants or data spaces. The distribution of this data among network nodes ensures independence from a central system. Consequently, each participating node has the ability to explore and manage participants.

Each node maintains the ledger and is updated as data changes. Updates are made individually for each node, and all nodes have the same level of authority after parties are certified and authorized as a data space authority.

### **Publication & Marketplaces**

As mentioned earlier, marketplaces, data sources and services can be linked under defined conditions.

This building block supports the publication of these offers, the management of processes related to the creation and monitoring of smart contracts (which clearly describe the rights and obligations for data and service use), and access to data and services.

### **For example**

A data space user requests the publishing platform for access to data sources of specific data assets. After selecting the dataset they want to access, they will receive a link with access to this dataset.

### **Data Usage Accounting**

This building block provides access to the usage data of individual users (logging). This, in turn, supports the important clearing, payment and billing functions (including data sharing transactions without the involvement of data marketplaces).

As an example, the clearinghouse in the Smart Connected Supplier Network (SCSN) is used to send purchase-to-pay information in a B2B scenario. This information can be highly confidential and is crucial to their day-to-day operations. If a disagreement arises, the clearinghouse is used as a third party to resolve the issue by comparing the fingerprint of the message to identify any errors.

## 6.4 Data Space Governance

### Overarching Cooperation Agreement

Chapter 5 explains the basis of the cooperation agreement, which covers all agreements in the data space per function block to ensure a federated exchange of data. These agreements are documented in the data space agreement, which may be officially signed during the onboarding process.

It is important to allocate a space for all elements from the 12 building blocks of the soft infrastructure.

The options for various blocks are detailed in the preceding chapters. Specifically, the cooperation agreement addresses matters related to trust and responsibility as well.

The iSHARE Legal Framework covers the basis of trust: the liability for violation of the use licenses and the primary confidentiality of data transactions.

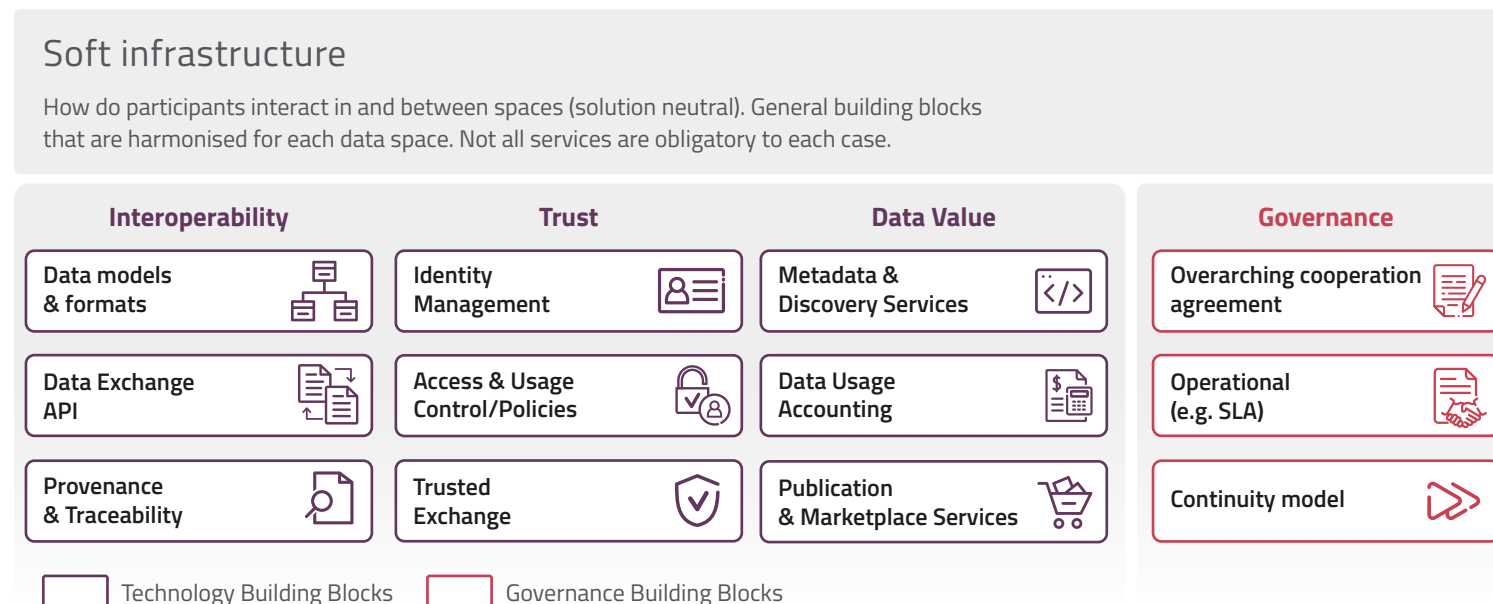


Figure 7: Data spaces building blocks, OpenDEI

Moreover, data spaces have the flexibility to include additional clauses, as exemplified below (though not limited to these examples):

- Financial contributions to the data space in general.
- Collective choices regarding the addition of new participants in the data space.
- Collective authorizations for, for example, data checks to maintain quality.
- Exceptions for specific players who have a different liability or role in the data space.

## Operational

The operational aspects concern all day-to-day elements of the building blocks, and the choices are extensive, but a few examples are;

- Data Exchange API: Ensuring guaranteed uptime, availability, and response speed of the servers.
- Governance: Establishing how monitoring of compliance with the rules in the data space is conducted (audits, frequency, reporting, etc.).
- Provenance: Determining what tracking the parties keep for proof from the APIs.
- Data Exchange API: Specifying which data services are available at least from parties.

## Continuity model

The final, yet crucial, aspect of data space continuity involves several key elements:

- Who monitors compliance in the data space to ensure its continuous operation?
- Who pays for which services in the data space? Is there a membership for all participants, and what value is provided in return?
- What rates apply to participation?
- What rates are there for the different data services?
- How does the admission process of new participants proceed to ensure interests are harmed?

# How to start with **Data Spaces and Building Blocks**

The initial step for nearly every data space involves forming a pioneering team or consortium to establish the inaugural application within a data ecosystem or data space. Ultimately, the objective is to work together and progress collectively towards achieving enhanced intelligence and excellence.

## 7.1 Bringing together initial Stakeholders

Data spaces unlock previously inaccessible data sources, enabling us to address significant challenges within the economy. The initiation of data spaces often arises from such shared challenges, fostering a sense of unity and a common objective. The parties involved are driven by intrinsic motivation, aligning their goals and priorities. Financial considerations are not necessarily the primary focus. When parties are primarily motivated by financial gain, there is a high likelihood of misplaced incentives. In the case of the most successful applications, financial gain typically ranks as a secondary or even tertiary driver.

Indeed, the pursuit of a win-win-win scenario, where all parties involved benefit or are appropriately rewarded for data sharing, is crucial. To promote the adoption of data spaces, it may be advantageous to engage influential

entities capable of steering ecosystems due to their substantial market presence, such as major players with numerous suppliers or governmental bodies. Their involvement can contribute to establishing a solid foundation and driving the success of data spaces.

## 7.2 The first federated application

The optimal beginning entails an application that possesses the following qualities:

1. Business sensitive data as a key component, as it allows the arrangement to be immediately ready for handling vulnerable data.
2. A market of software vendors, where there are multiple market leaders with large amounts of data from a large number of players.

3. A target which no one can oppose and which benefits everyone in the data space. The UN Sustainable Development Goals provide a fantastic source for themes, especially since we have a desire to make a positive impact on the world and take meaningful action.
4. The first Data Service is preferably limited, not a service with hundreds of fields, but a limited set or something that already exists. This lowers the threshold.

Good examples of practical applications in line with these points are:

- CO<sub>2</sub> insights reports for transports
- Energy-saving reports for homes or businesses
- ETD (Estimated Time of Departure) insights

More examples will be mentioned later in Chapter 9.

### 7.3 Start with contracts and legal coverage of the exchange

Many experiments overlook crucial aspects, leading numerous data spaces to become excessively burdened by delays due to the realization that separate Non-Disclosure Agreements (NDAs) still need to be signed with each party involved, often at the eleventh hour.

There is a need for legal coverage of the exchange of data in a number of areas, as mentioned earlier in chapter 5.4, but there is a scale per area.

#### Data Licenses

The foundation lies in providing a clear and definite understanding of how the data will be utilized after its reception. Using a non-disclosure agreement may not be suitable in this situation since there could be a particular objective for each data exchange, which could make sharing necessary or beneficial. That is why legally covering user licenses, and signing a digital contract for every data interaction, is the proper route to quickly start with sufficient certainty.

This encompasses not only the reading of data but also the writing of data, which necessitates specific licenses. Neglecting this crucial aspect would undermine the fundamental purpose of data exchange.

#### Data Reliability

Quality of the data is certainly crucial, but this can differ greatly per data space, for example, if impactful decisions are taken based on the data.

## Data Services Availability

The availability (how fast a service responds, how fast is the data connection to a service, how often there can be disruptions) is important and must be determined for the application within the data space.

## Data Services Security

In addition to the security of the interaction between the data consumer and the data provider, the security of the servers, services and applications within the data space is important. In principle, liability is already a sufficient incentive to deal with carefully, but it may be desirable to add an extra element.

## Generic cross data space Legal Framework versus Proprietary and Closed

You have the option to either develop your own comprehensive legal framework from scratch or utilize a generic foundation and enhance it with additional verifiable clauses.

The latter is how iSHARE works, a generic framework for all data spaces that provides a basis in terms of licensing, availability, and reliability, and can be supplemented with extra clauses. The extra clauses give the freedom to create conditions that are specific to each data space.

This approach ensures legal and functional interoperability between data spaces and allows data sources to be queried at the same levels of certainty across data spaces.

The iSHARE Framework is generic, with governance for further development.

## Start with the legal basis of iSHARE

Starting with iSHARE's generic legal basis is easy, saves time and legal costs. This basis gives complete freedom to expand later and develop new clauses and ensures connection with other data spaces. After all, all parties within data spaces based on the iSHARE Framework have certainty about the legal coverage of the shared data and the acceptance of the associated licenses.

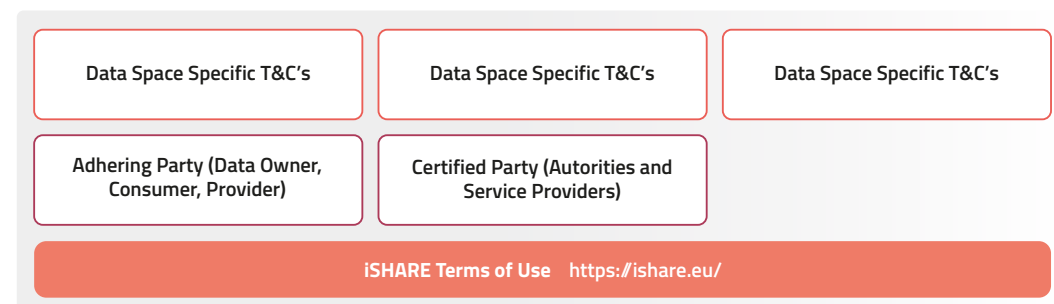


Figure 8: iSHARE Terms of Use, iSHARE



## 7.4 Registering the companies in the ecosystem and setting up authorizations

Another element that often lags behind in many processes is not just the registration of parties, but the implementation of digitally verifiable methods. Often, the initial focus is on establishing technical connections and facilitating data exchange between suppliers and users. While this accelerates progress, it can lead to a loss of oversight regarding signed contracts, the integrity of certificates, and the secure utilization of tokens.

The iSHARE Satellite provides the fast and future-proof route for this. This trust anchor provides a validation interface for all players to verify whether parties are legally, administratively and technically correct and compliant with the agreements that apply within the data space.

At the same time, it informs the basis for granting authorizations/permissions. After all, you can authorize someone else if you are certain that the other party can be trusted.

Using iSHARE as a basic registration environment ensures that all organizations are interoperable and discoverable for other data spaces. Instead of being a risk, it presents an opportunity for mutual learning and collaboration among participants, based on equal legal grounds.

## 7.5 Unlocking the first Data Suppliers

Common objections raised by data providers regarding this issue include:

1. "That can be done through email" (commonly heard in logistics, for instance).
2. "That can be accomplished with a dedicated link" (as it aligns with their business model).
3. "You can easily download it through our application, making it simpler."
4. "But that's my data, and I can't freely share it without restrictions."
5. "My customers are concerned about the potential risks and consider it dangerous."

The purpose of this chapter is to proactively address and overcome these objections, ensuring the continuous progress of the data sharing project. Its focus is to elucidate the value proposition, particularly for the data providers, in order to sustain momentum.

The first crucial factor is to ensure that a clear customer demand/need is felt by, for example, suppliers of a major customer or a major government party.

This is followed by responses to the objections raised by parties who view data spaces as complex, overly complex, or unnecessary:

1. Yes, but there must be a contract with all parties receiving the email, and that cannot be validated with email. Furthermore, this is manual, so there is a chance that the wrong message will be sent to the wrong person, and you have no control over what happens to the message afterward.
2. That's right, but then we're going to make 100,000 links, now we aim for 1 link for all players, which also keeps your software more manageable.
3. Yes, that is indeed the case; however, I still lack control over how my data is used by others.
4. Exactly, so with this you can determine under which conditions you share which data and therefore not share it.
5. All current manual procedures are much less secure than the data spaces discussed.

When this phase is over, the real process of connecting the data provider begins, and it takes place in a number of steps:

1. Signing the terms of use and the onboarding procedure (Ascension Agreement).
2. Deploying the Connector, (e.g., iSHARE Connector (IDS 1) or extended connector IDS 2 or 3).
3. Deploying the data service (e.g., based on a Linked Data standard) and a small service that covers the basic needs of the application.
4. Testing the connector implementation using a test bed, conformance testing tool, or other approach.

During this phase, the involvement of other roles is also essential. We will explain the practical first steps to set up these roles in chapter 10.

## 7.6 Acceleration Programs

Several programs, such as i4Trust.org, DIL (Data Infrastructure Layer), Data Spaces Support Center, and the Adoption Support Center, offer valuable assistance in terms of accelerating progress and providing clarification.

# Interoperability of Data Spaces

Throughout this book, we have repeatedly emphasized the significance of interoperability among data spaces.

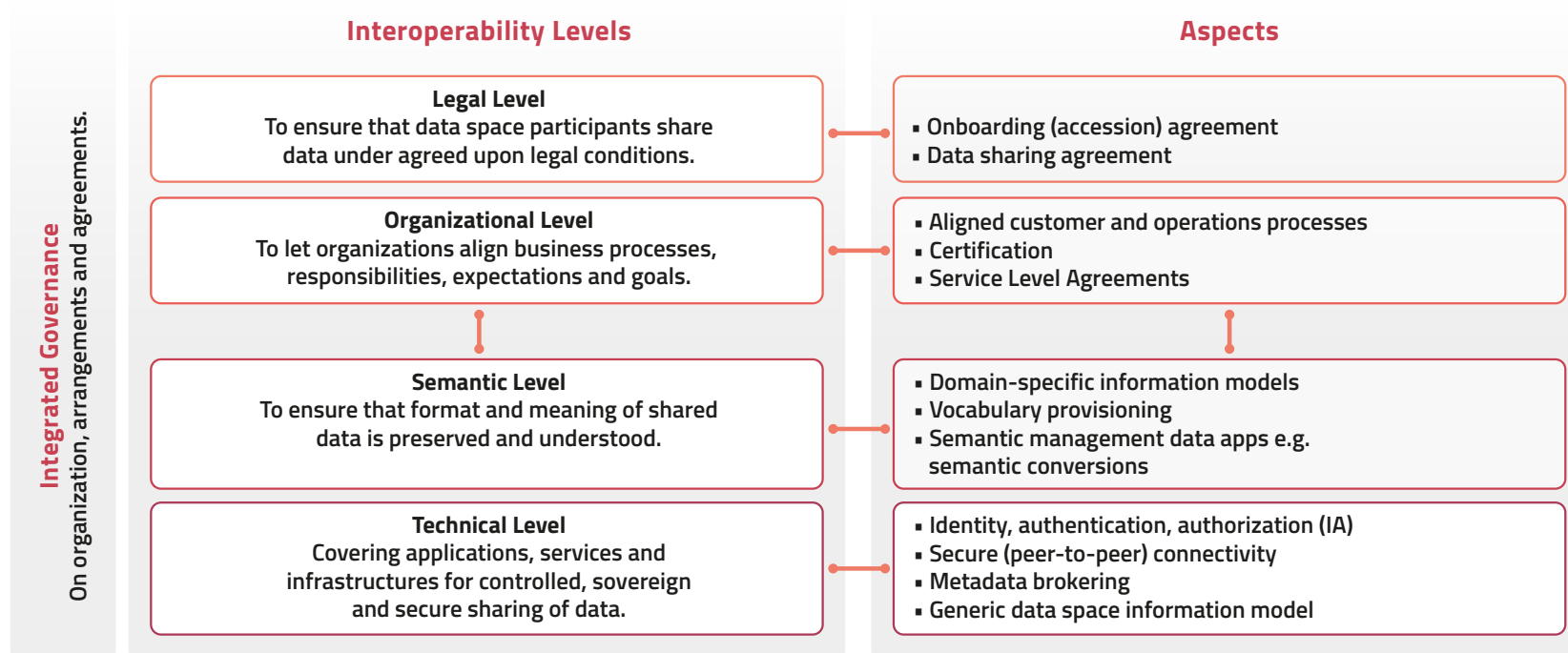


Figure 9: Interoperability Framework European Commission, EIF

Data space interoperability is more than just the interoperability of technical components. The European Interoperability Framework (EIF) developed by the European Commission, provides an approach to systematically categorize interoperability aspects. There are four levels of interoperability that fall under an overarching integrated governance approach:

- Legal level
- Organizational level
- Semantic level
- Technical level

#### Legal level

At the legal level, the focus is on ensuring that organizations operating under diverse legal frameworks and strategies can effectively collaborate and function together.

This entails ensuring that legislation does not impede the development of European public services, both within individual Member States and across borders. It requires establishing clear agreements on how to navigate disparities in legislation and explore avenues for introducing new regulations when necessary.

Legal interoperability can be addressed by conducting interoperability checks, screening existing legislation to identify interoperability barriers. This includes sectoral or geographic restrictions on the use and storage of data and different and vague data licensing models.

To ensure interoperability, coherence between legislations should be assessed before adoption and through regular evaluation of performance once they are adopted. Given that European public services are increasingly delivered through digital channels, it is crucial to consider information and communication technology (ICT) from the earliest stages of the legislative process. This proactive approach ensures that the legislative framework aligns effectively with the digital nature of public services.

#### Organizational level

At the organizational level, the focus lies in aligning all business processes, responsibilities, expectations, and goals. This ensures cohesion and harmony within the organization, enabling smooth operations and shared objectives.

### Semantic level

Semantic interoperability ensures that the precise format and meaning of the data and information are preserved and also understood during the data exchange. Within the European Interoperability Framework (EIF), semantic interoperability encompasses both semantic and syntactic dimensions, including:

- The semantic aspect refers to the meaning of the data elements and the relationship between them. This involves creating a common language to ensure that all parties involved in communication share the same understanding.
- The syntactic aspect refers to the description of the exact format of the information to be exchanged in terms of grammar and format. In Data Spaces specifically, the principle of Linked Data has been developed to ensure that it can always be traced back to a source meaning.

### Technical level

At a technical level, achieving interoperability involves the integration of applications and infrastructure, which act as the connecting components for various systems and services. Technical interoperability encompasses crucial elements such as interface specifications, data presentation and exchange methods, and the implementation of secure communication protocols.

A major obstacle to interoperability stems from legacy systems. Historically, applications and information systems in government services, for example, have been developed bottom-up to solve domain-specific and local problems. This led to divided IT islands that find it difficult to work together.

Due to the size of the public administration and the distributed ICT solutions, the abundance of legacy systems creates an additional interoperability barrier in the technical layer. Technical interoperability must be ensured through formal technical specifications.

Examples of the value of this interoperability can be found, for instance, between the Dutch Data Space for Energy (BAS), data space Sustainability Utilities (further explained in chapter 10), and Digital System Built Environment. Consider the example of building management where energy consumption is a good predictor of problems in the building where a smart meter is already installed.

This energy consumption does not say anything without clarity about the properties of the building. DVU has opened up the historical data of buildings within that system, BAS has combined the data from smart meters with the owner of a building.

In any case, there is a lot of potential for new services by linking the various domains.

# Data Spaces **examples**

Fully realized data spaces with federated applications, as intended within the context of data spaces, are relatively scarce. While many researchers have explored this concept, practical implementations often vary in subtle ways. In this cookbook, we have compiled numerous practical lessons learned from real-world experiences. Consequently, this chapter does not present theoretical examples but instead focuses on practical illustrations that highlight the key lessons derived from these real-world cases.

## 9.1 Data Space for Sustainability of Non-Residential Buildings (DVU)

Buildings must be made more energy efficient to achieve the objectives of the Climate Agreement.

The aim of this data space is to provide faster access and better insights into the actual energy consumption in the business premises in the Netherlands. The data space allows structural comparison of energy consumption per m2 possible for different types of buildings. It is also being developed to make this data more accessible and to share it in a secure manner.

DVU is fully federated in terms of interface standards but starts with a combination of data reporting insights,

Authorization register and registration process. We explained this application in detail earlier in this cookbook. More information can be seen in the **demo** and in the **explanation video**.

Some of the lessons from this data space are:

- Using an identity tool such as E-Recognition makes the onboarding of data owners much simpler.
- Data sets can be labeled and expanded directly in the registration of data owners. In this case: this is my business premises, this is the configuration of the premises, etc.
- Creating policies for different applications also requires a different way of registering and managing, so the user interface of an Authorization register could very well be different per data space. In addition to an

Authorization register policy control interface, there is a need for a definition of an authorization creation interface for this type of application.

## 9.2 Smart Connected Supplier Network

Collaboration within the supply chain is becoming increasingly vital, necessitating the sharing of critical data such as orders, logistics information, and technical data. However, the transmission, reception, and processing of this data can be expensive and prone to errors. The Smart Connected Supplier Network (SCSN) serves as a data standard that effectively streamlines this process, facilitating efficient and accurate data exchange within the supply chain.

The SCSN makes the exchange of information in the supply chain more efficient, allowing companies to share data in a secure and interoperable way. This, in turn, results in higher productivity of the supply chain.

Some lessons from this data space:

- SCSN has successfully leveraged the platforms of the leading software providers in the market, establishing a network with five key software players that collectively serve a vast number of parties.

- When defining interface standards (APIs), it is crucial to strike a balance and avoid excessive complexity. The API specifications should be designed in a way that allows software players to effectively implement them.
- Governance of the participants requires good procedures and structure across the players, in which the SCSN foundation plays a trusted role.

## 9.3 Catena-X

Catena-X represents the pioneering integrated, collaborative, and open data ecosystem specifically designed for the future of the automotive industry. It serves as a comprehensive platform that connects all stakeholders across end-to-end value chains. Catena-X employs cutting-edge technology to establish an open and enterprise-wide system. By enabling data providers to retain control, Catena-X empowers them to determine the participants, timing, and manner of data exchange, ensuring data sharing occurs on their terms.

Catena-X has the Eclipse Data Space Connector (EDC) as its central communication component, which forms the connection for sovereign and cross-organizational data exchange.

The guiding principles in the development of the EDC are independence, simplicity and the maintenance of a small and efficient core. The required functionalities are bound together in the open source project “Eclipse Datas Space Connectors”, to which the Catena-X partners contribute.

Some lessons from this data space:

- The definition of first scalable and federated services is a good place to start. For example, Catena-X launched the Federated ETA as the first service to ensure that all parties were the first to use it.
- The process of establishing new legal and governance frameworks can be more demanding than initially anticipated. However, implementing a structure such as iSHARE can be immensely beneficial in this regard.

## 9.4 DASLOGIS

The DASLOGIS project is dedicated to creating and showcasing the Dutch Logistics Data Space (DLDS), a data space specifically designed for the Dutch logistics sector. The primary objective of this initiative is to enable organizations within the ecosystem to seamlessly and securely share their data while maintaining data sovereignty.

The project is funded by TKI Dynalog as a research project to understand and test the reference architecture of International Data Spaces in a proof of concept setting for the Dutch logistics sector.

DLDS relies on the generic and internationally standardized IDS Reference Architecture Model, developed by IDSA (International Data Spaces Association). The organizations involved, including iSHARE, test the technical feasibility and business value of a DLDS on the basis of logistics use cases.

Some lessons from this data space include:

- In many instances, participants may lack a clear understanding of terms such as “federated” and “data space.” Thus, providing a precise definition and clarification of the question becomes necessary.
- In this context, a single federated data service, namely the Federated ETA, was selected to ensure that all parties are aligned and focused on the same objective.
- Fully digital usage control is still difficult to implement without certified apps and connectors. In such cases, leveraging iSHARE’s legal usage control can serve as a valuable initial measure.



## 9.5 Statistical Data Space NL

In the Netherlands, there are several data spaces that generate statistical data reports for the Central Bureau of Statistics (CBS). A few examples of these are Modality reports and Vehicle Emission reports:

### Modality Reports for Rijkswaterstaat

One of the statistical challenges facing Rijkswaterstaat is how to collect sufficient correct data with the smartest possible structure. For Rijkswaterstaat, a crucial factor is determining which transport infrastructure is used for the transportation of specific quantities of goods. This can be done by having CBS collect the data from all carriers, but it can also be done by requesting authorization for CBS to access this data through other sources.

Together with CBS, Poort8 and Modality software, a complete chain has been set up for collecting modality insights under the authorization of the data owner (the inland shipping terminals and possibly even the shippers).

By virtue of being the software supplier for 23 terminals, Modality has the capability to consolidate data streams from multiple terminals into a single link. This approach not only ensures efficiency but also establishes a secure and well-structured data space. The data owners, i.e.,

the transporters, explicitly authorize the sharing of their data through the implementation of iSHARE, thereby creating a trusted environment. As a result, accessing data within this data space proves to be more cost-effective compared to alternative options.

A lesson from this data space:

- Unlocking data from a more central location for many organizations at once saves a lot of costs and immediately creates significant value.

### Vehicle Emission Reports

Data from the logistics sector provides crucial insights into the state of the economy. The Vehicle Emission Shipment Data Interface (VESDI) is a project designed to provide an easier way for transport companies to submit their basic data to Statistics Netherlands.

Within this sector, there is a great willingness to share data, but they often encounter obstacles due to the considerable attention, time and money required for enabling automated electronic delivery. As a result, an electronic interface has been developed based on the Open Trip Model (OTM), an open standard established through cooperation between logistics service providers, IT suppliers and shippers.

The project's primary focus is on sharing 'basic' data about the vehicles and the goods. It aims to indicate the means of transport, registration number, type of goods, and route used for transportation.

Lessons from this data space, where authorizations are not required, and data is sent or uploaded at CBS:

- The digital maturity of the parties differs enormously, necessitating alternative methods of data exchange when direct links are not feasible.
- The data standard and trust are also important in this context, and there is room for improvement in organizing them more uniformly rather than bilaterally.

## 9.6 I4Trust Projects

i4Trust is an initiative of iSHARE, FIWARE and Fundingbox aimed at accelerating data space development in Europe through an acceleration program that offers funding, technical support and mentoring. The program has launched 13 data spaces, with another 17 data spaces currently being launched. Two appealing examples of these projects are CO<sub>2</sub>-Mute and eVine2Wine.

### CO<sub>2</sub>-Mute

CO<sub>2</sub>-Mute supports the use of alternative mobility by proposing achievements to commuters and by providing insights to public authorities on sustainable mobility policies.

The transport sector is one of the largest contributors to environmental impacts due to greenhouse gas and CO<sub>2</sub> emissions.

The CO<sub>2</sub>-mute project aims to support local governments to involve citizens in implementing policies for sustainable mobility and urban green infrastructure.

CO<sub>2</sub>-Mute has a high degree of flexibility in configuration and also by being "interoperable by design" has a great ability to handle different types of data. Also by using a Digital Twin platform implementation based on the NGSI-LD standard.

The ability to personalize and correlate the data, based on the local situation, is crucial to engage citizens in concrete actions that lead to measurable impacts. The safe and controlled sharing of data plays a central role, as business models are needed and CO<sub>2</sub>-Mute relies on the i4Trust architecture.

## eVine2Wine

eVine2Wine shares relevant grape production data from the vineyards along the value chain, aiming to provide a superior product and customer experience.

Currently, traceability from wine to vineyard (chain traceability) is a desirable and much-discussed topic within the wine industry, mainly for two reasons:

- For quality wine as a premium product, such information holds value for marketing and sales purposes.
- To be able to provide a reliable overview of the traceable wine production conditions and the past activities that may have an effect on the safety and quality of the wine.

The concept of fully sharing vineyard data and traceability is currently only being used by a limited number of wine producers. With today's technology, this can be taken to a higher level.

The aim of eVine2Wine is to provide reliable and controlled data exchange using Digital Twins of the vineyard areas, with iSHARE providing identity and access control mechanisms. For instance, a vineyard manager will then be able to determine which data from their vineyard knowledge database may be shared with the wine broker to whom they supply grapes.

## Lessons from i4Trust

In i4Trust, we have seen that it also requires some effort for software developers to understand what "federated" means and that trusting external identities or external authorizations can indeed work well.

To make this concept easy to understand, i4Trust offers a set of open-source components and accompanying mentoring. Thanks to these components, software developers can quickly take steps in testing and experimenting with federated components. More information about this can be found on the i4Trust [Github](#).

# Get started step by step with Data spaces

## 10.1 The optimal start for “Data Provider”

Developing federated services in a data space can be daunting or complicated for many parties. It is wise to commence with a foundation where parties already possess a fundamental interest in sharing information, avoiding complexities associated with financial transactions, individual integrations, and potential vulnerabilities. By starting from this standpoint, the process of information sharing can be streamlined and focused on fostering collaboration and mutual benefits without unnecessary complications.

It is logical to start with regular (i.e. recurring) data exchange based on reliable sources, such as on CO<sub>2</sub>, statistical information, government obligations, and energy savings. When we refer to reliable sources, we are not advocating for separate data services per company, where data quality needs to be individually validated. Instead, we emphasize the utilization of trusted sources such as an energy company's data, land registry information, or a port community platform. These established and authoritative sources offer a higher level of data reliability and accuracy,

eliminating the need for redundant validation processes and ensuring consistent data quality throughout the ecosystem. Parties that have their data quality in order from home because it forms the basis of their daily work.

A second crucial factor for rapid adoption is avoiding 1-to-1 and aiming for n-to-n connections. Start with data providers that have datasets from as many companies as possible. This refers to data providers such as SaaS (Software as a Service) suppliers, service providers (Telco, Energy, Building management), and central service providers (Land Registry, CBS etc.).

The third factor is simplifying/keeping the first data service simple. This service must be recognizable in itself, be ready for demand, and have multiple possible applications. But not too extensive, starting with 7-10 data fields in the APIs, so that you can start with a simple service.

Ideally, it starts with an existing API from the service provider that is accessed on the basis of the identification, authentication and authorization standard in the data space. This enables rapid implementation without too many adjustments.

A crucial factor for the identification of the different parties within the data space is the labeling of the data to the data owner. An identifier such as the EORI number or another ID of the different parties must be available to the API as a filter. Consider the earlier mentioned energy data, which provides information on the energy consumption of a specific address. However, it is crucial to determine the corresponding company occupying that address and, consequently, serving as the data owner for that specific data. This process of assigning the data to the appropriate data owner is known as data labeling, which ensures clear attribution and accountability for the data associated with each entity.

In DVU, for example, this has been done by launching an extra service that bridges the gap between EAN (Energy connection) and EORI (company identity) of the data owner/contractor of that connection.

## 10.2 De optimale start als “Data Gebruiker”

Because trust is a basic component at the beginning of every data space, it is advisable to begin by engaging a generic trusted Data User as the initial step. While subsequent steps may involve different Data Users, this initial phase is pivotal in building the foundation of trust within the data space ecosystem.

Examples of “trusted Data Users” include reputable government entities and agencies like CBS (Central Bureau of Statistics), RVO (Netherlands Enterprise Agency), Inspectorate, and NVWA (Netherlands Food and Consumer Product Safety Authority). Additionally, established players in the data space, such as industry associations, can also serve as trusted Data Users. Furthermore, parties with whom you have a longstanding relationship as a Data Provider can be considered as trusted Data Users, given the established trust and track record of collaboration over the years.

The Purpose/License of the first application of the first Data User must not be threatening but immediately offer time savings for the data owner, such as:

- This data is only used for anonymous reports that cannot be traced back to the company and not for commercial purposes (iSHARE License 0003)
- This data is only used to determine energy savings for the data user and is not shared with third parties. (iSHARE License 0002, Internal Use Only)

This makes it clear to parties that savings in reporting time do not entail any further risks.

## 10.3 The optimal start of the Data Owners

Ideally, we initiate the data space by involving a small group of influential data owners who can serve as catalysts for engaging the broader community of owners. These key stakeholders hold significant influence and are respected within their respective domains, making them effective connectors to the data space.

## 10.4 First start of the Data Space

To start practically, the first steps are to set up the basic data space elements. This is explained using the iSHARE governance and technical components.

### Governance & Legal

When considering the establishment of governance for a data space, there are two options: setting up your own governance framework or joining existing trust frameworks like iSHARE. For parties seeking a prompt start, joining an existing trust framework is recommended. This choice involves determining the legal foundation and the required level of trust within the data space. The level of trust is determined by the highest level needed for the most demanding application within the data space.

In the case of the iSHARE trust framework, the necessary actions are as follows: register the (digital) signed contracts and/or clauses and turn them into digital verifiable credentials in the iSHARE network via the iSHARE Satellite.

Test and certify both the data provider and the data consumer on a technical level and register this in the iSHARE network.

### Technical

To implement the first federated application, a number of functions (roles) are needed in the data space, as described in chapter 4. The following components are available to set up these roles in the simplest version:

#### Trust Anchor

Deploy an iSHARE Satellite using the manual at <https://github.com/iSHAREScheme/iSHARESatellite>

#### Authorization Register

Start with an Authorization registry with the open-source version of iSHARE available at <https://github.com/iSHAREScheme/AuthorizationRegistry>, or select an existing supplier of turnkey Authorization register services.

### iSHARE Service Provider

Implement the iSHARE service provider based on the open-source version at <https://github.com/iSHAREScheme/ServiceProvider>

### Service Consumer

Deploy the first Service consumer and use the documentation available at <https://github.com/iSHAREScheme/Documentation>

Further documentation available at <https://dev.ishare.eu>

## Operational

### Reliability APIs

For the operational aspects of the data space, the reliability of the APIs is important. Therefore, define the operational levels as core elements in the data space:

- Availability of APIs: e.g., 99.9%
- Quality of data: Semantics
- Reliability of the data: Timeliness of the data

Op <https://ishare.eu/benefits/for-data-spaces/> is veel meer te lezen, en zijn de basis Service Level Agreements in te zien die de basis vormen van iSHARE.

### Operational Processes in a Data Space

When setting up a data space, several practical aspects must be considered:

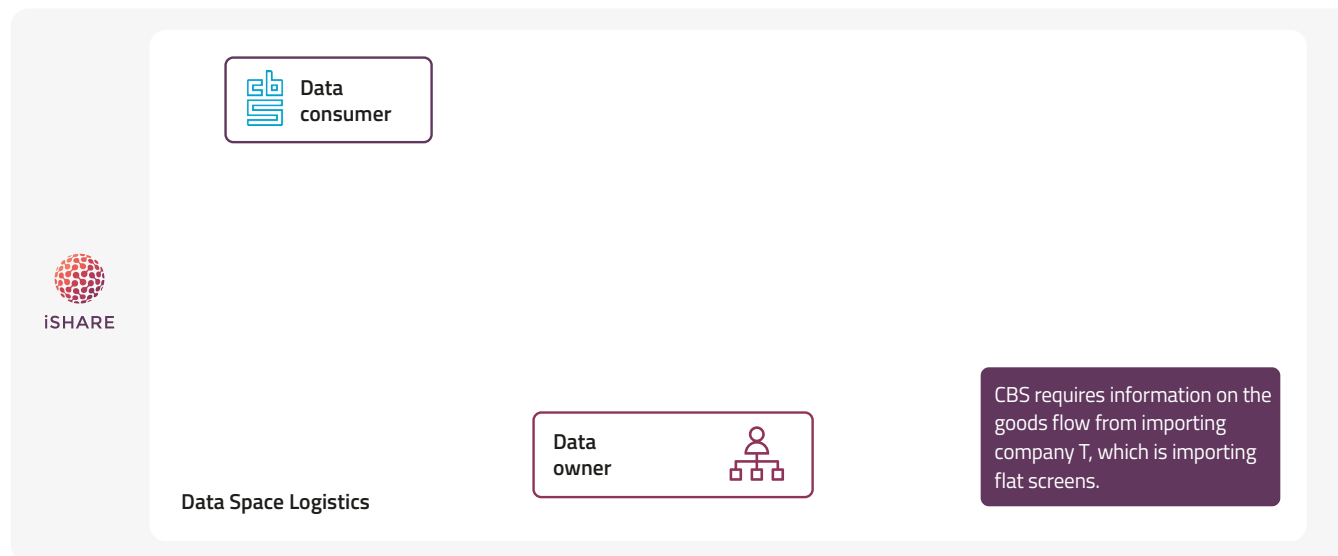
- Admission
- Withdrawal
- Warnings, Suspension, and Exclusion
- Incident Management
- Release Management
- Management Reporting

If the iSHARE Trust Framework is used, the iSHARE is built on **operational procedures**.

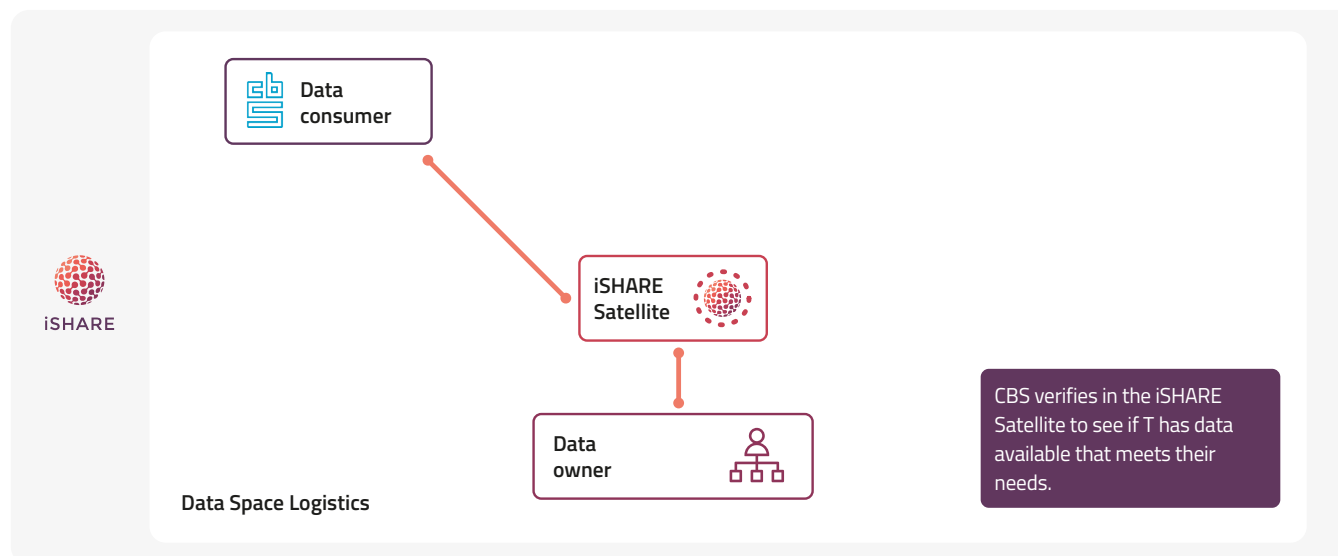
## 10.5 Pilot the first Data Space application based on the deployment

Once the technology has been implemented, it is time to take the initial steps to conduct testing. The schematic figures below illustrate an example case, outlining the key stages of the process.

Test the entire data access procedure.

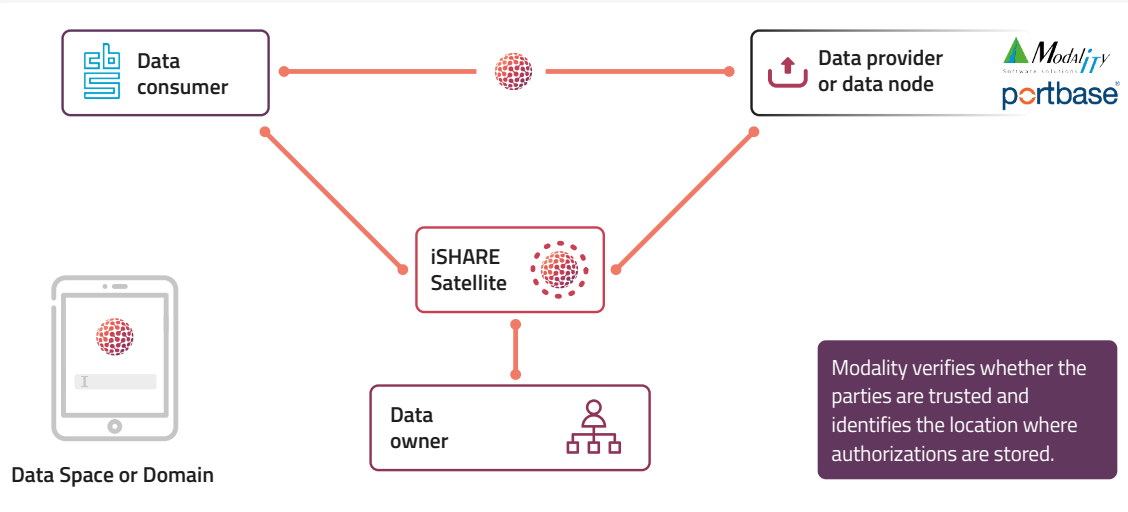


The Data Consumer requests the location of the data APIs of the owner in the satellite.

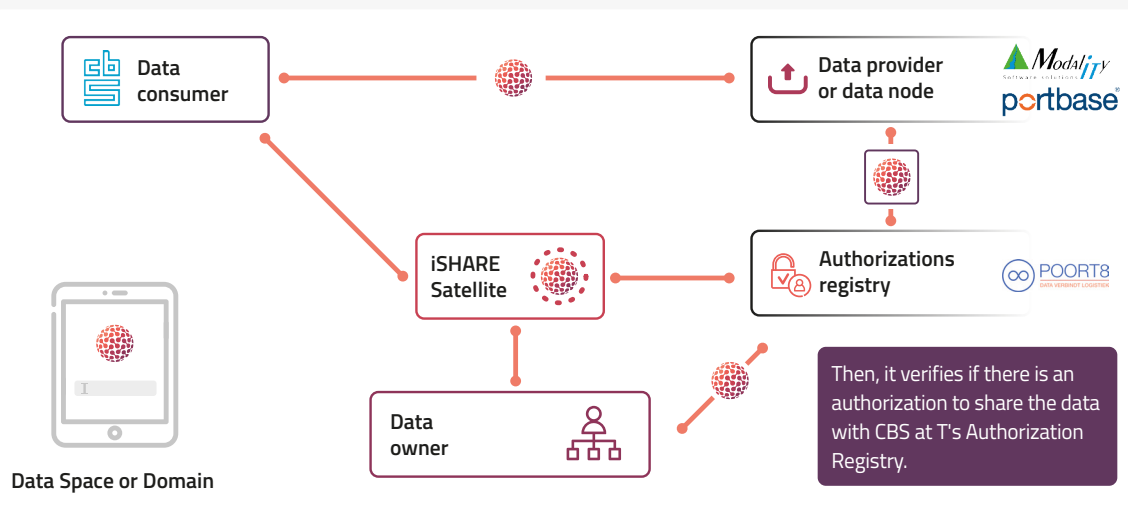


The data consumer requests authorization to the Data provider for a first owner.



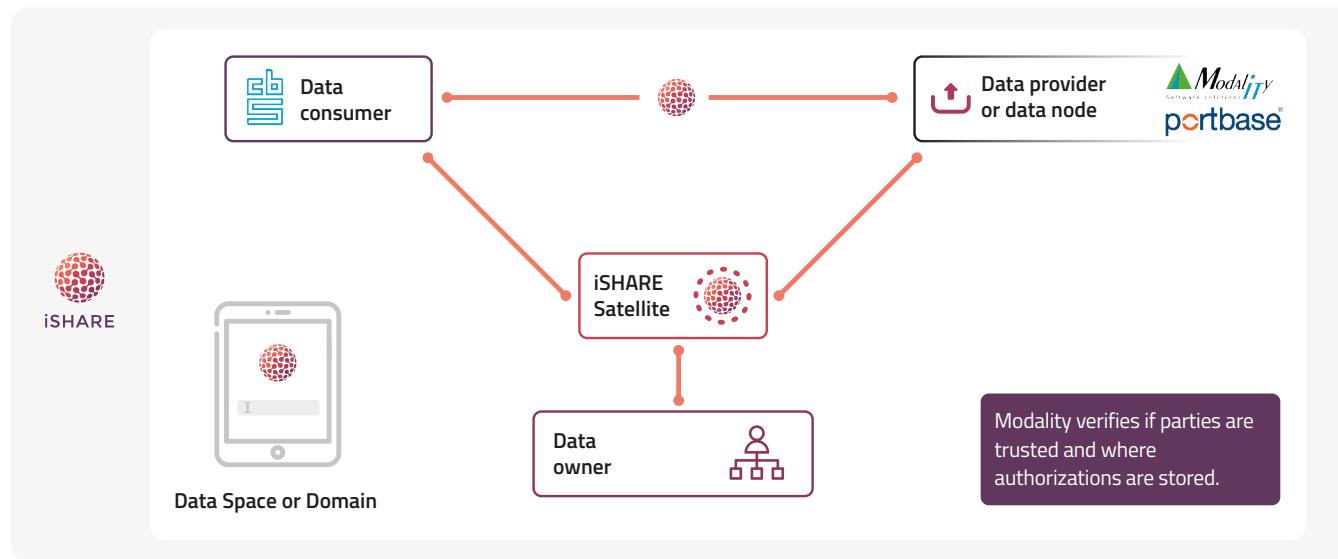


The data service provider performs permission checks against both the satellite/trust anchor and the authorization registry to ensure compliance.



The data owner receives the request for authorization and registers the Policy.





The Data consumer verifies again with the data data provider to ensure that the authorizations are now properly registered.


# Conclusion

Data Spaces, which involve agreements among ecosystems of companies with a shared objective of data exchange, offer numerous opportunities and possibilities. By allowing data owners to retain full control over their data while leveraging the services and expanding the data space ecosystem, exponential potential can be unlocked. This convergence of data ownership, services, and growth opens up new horizons for

## **innovation and collaboration.**

This book serves as a comprehensive overview of the latest developments, roles, and possibilities in the field. Our aim is to facilitate an accelerated shift in thinking among various stakeholders, providing readers with valuable insights to enhance their understanding and approach

The establishment of data space structures unlocks limitless possibilities to dismantle chains and silos, paving the way for smarter and more sustainable practices. With data as the driving force, this transformation encompasses the entire system rather than just a limited portion.



We firmly believe that adopting a systemic and value chain-oriented approach, fueled by data and data permissions, will be the catalyst for the innovations that will shape the next decades. This involves breaking down barriers and embracing collaboration across the entire ecosystem to drive meaningful change and unlock the full potential of data-driven innovation.

Therefore, we hope that this book has inspired you to take real action. To consider where data can be utilized, where there is potential value to create smarter chains with existing data. And to take the first step towards a better world by harnessing the power of all the data that already exists!



## Colophon

**Authors:** Gerard van der Hoeven & iSHARE Team

**Titel:** Cookbook for Data Spaces

Self-published | © 2023, iSHARE Foundation | [info@ishare.eu](mailto:info@ishare.eu)

All rights reserved.

No part of this publication may be reproduced, stored in an automated database and/or made public in any form or by any means, be it electronically, mechanically, by photocopying, recording or in any other way, without the prior written permission of the publisher.

