# Fostering Digital Transformation and Data Valorisation in EUSALP

## Enhancing Competitiveness and Innovation through Data Spaces

Interreg
Co-funded by the European Union
Alpine Space
Support EUSALP
EUSALP EU Strategy for the Alpine Region
REGIONE PIEMONTE
FONDAZIONE PIEMONTE INNOVA

Interreg Alpine Space Programme (alpine-space.eu)

**Support EUSALP** | Technical Support Structure for the Governance and Implementation of the European Strategy for the Alpine Region

Coordinated by: **Gianmarco Piola** – Piemonte Innova Foundation

Research by: **Giulia Boccaccio, Maria Montarsolo, Carolina Tardito** – Piemonte Innova Foundation

Graphic design by: **Marco Tattoli** – Piemonte Innova Foundation

Acknowledgements

We would like to extend our thanks for the interviews to:

**Caroline Lange** – COOPERANTS
**Davide Dalle Carbonare**, **Angelo Marguglio** – Circular TwAIN, Eur3ka
**Ulla Sparre** – Water Data Space (Water Valley Denmark)
**Gerasimos Antzoulatos**, **Roberto di Bernardo** – WATERVERSE
**Gian Antonio Battistel**, **Nicola Cracchi Bianchi**, **Luca Giovannini** – USAGE
**Valentina Caron**, **Davide Coslovich**, **Vincenzo di Felice**, **Lara Manzoni**, **Stefano Predan**, **Valentino Volpe** – AgriCS
**Fritz Fahringer**, **Raphael Schaller** – datahub.tirol
**Robin Bauer**, **Andreas Schmitt** – Truzzt Alliance

We sincerely appreciate the invaluable contributions of the following individuals to our study.

**Stefano Modafferi** – Big Data Value Association
**Sophie Meszaros** – European Data Space for Smart Communities
**Roberto Monsorno** – Eurac Research

Publication date: **May 2025**

Notice: This publication is part of the INTERREG Alpine Space project Support EUSALP.
The project is co-financed by the European Union via the European Regional Development Fund.

The authors are responsible for the content of this publication.

# ABSTRACT

The European Data Economy involves the collection, processing, sharing, and monetization of data across diverse sectors and industries, with the goal of creating value, driving digital transformation, and cultivating a data-driven society. The macroregional strategy EUSALP aims to work in synergy with these objectives, promoting and supporting these principles in its initiatives.

This study therefore examines the **conditions for the development of a data sharing ecosystem in the Alpine region**, orienting the valorisation of information assets towards the public good, territorial cooperation, economic growth and sustainability.

Based on an analysis of established and emerging models for data sharing, governance and valorisation, in-depth research on data spaces is proposed as a new approach to bring together relevant data infrastructures and governance frameworks to facilitate data pooling and sharing. The work highlights the strategic role of data for strengthening administrative capacities and for building innovative, interoperable and adaptive services to local needs.

Supported at European level, the adoption of data spaces represents a concrete opportunity to promote new forms of collaboration between public institutions, private actors and research centres also in the context of the Alpine arc. For this reason, the study delves into some experiences and initiatives developed in relation to two themes of strategic importance for the EUSALP area, namely manufacturing as a service and water management.

The present work aims to show that **a data space is not solely a technology platform or tool that allows authorities to exchange data, nor a simple catalogue of data but is guided by several design principles related to data control, governance, respect for European rules and values, technical data infrastructure, interconnection and interoperability, and openness**. For this reason, the research concludes with some policy recommendations to promote public-private dialogue on data spaces as an approach that can foster the growth of the entire Alpine region ecosystem.

Within this framework, policy tools are discussed that are intended to experiment with multi-level governance solutions that strengthen the decision-making capacity and digital autonomy of local institutions and make each enterprise part of a generative ecosystem. **The Alpine arc is thus configured as a privileged context for developing practices of digital sovereignty capable of combining economic development, innovation and territorial cohesion**.

# Sommario

# Data Valorisation Strategies

## From Storage to Data-Driven Business Models

Interreg Alpine Space

Co-funded by the European Union

Support EUSALP

EUSALP EU STRATEGY FOR THE ALPINE REGION

# CREATING VALUE FROM DATA

The growing importance of data requires both private and public actors to find the best ways to leverage the data they all produce, directly or indirectly, during their activities. The following pages attempt to trace a path that, starting from an explanation of different models and architectures for data storage, will show some examples of data-driven business models, made possible thanks to the integration of data from different sources and the synergy between the business world, research, and public administration in offering to end users services that are increasingly comprehensive, effective, and responsive to real needs.

## Data Storage

Among the foundational actions that an organization – whether a private enterprise or a public institution – can undertake is the systematic and proper storage of data and datasets. A key distinction must be made between leaving data dispersed across multiple, often uncoordinated repositories, and adopting a centralized data storage system. The latter enables organizations to actively leverage their data assets: it ensures **visibility** – by providing awareness of what data is available – and **accessibility** – by making it possible to locate, retrieve, and utilise data efficiently and for multiple purposes. Storage architectures are structures and methodologies used not only to store and protect data but also to analyse them and obtain new information - and thus new value, bringing companies that strategically use them progressively closer to data-driven business models.

### Data Warehouse

Data warehouses[1] are **centralized repositories that store and harmonize large amounts of structured and semi-structured data**, which are analysed for business intelligence purposes, reporting, data visualization, and - more generally - with the strategic goal of enabling more informed decisions. Within these architectures, **data from various sources** (such as transactional systems or relational databases) **are stored at regular intervals** (not in real-time) **and standardized to allow integration**: data warehouses are relational in nature, so the structure and schema of the data are predetermined by the company's and product's requirements and optimized for SQL query operations. The goal is not to minimize redundancy through normalization but to store data in a way that favours the production of information through read-only access (the so-called "schema-on-read").

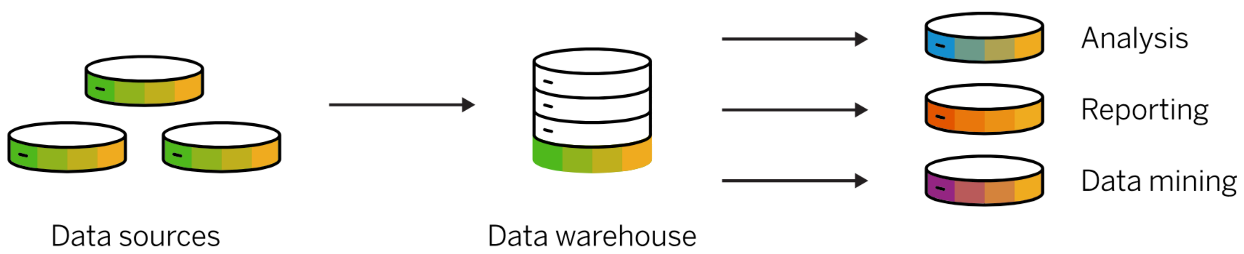



*Figure 1. What is a data warehouse. Source: SAP*

[1] Sources for this section:
What is a Data Warehouse? Data Lake vs. Warehouse | Microsoft Azure
What is a Data Warehouse? - AWS
What is a Data Warehouse? | IBM
What is a data warehouse? | Definition, components, architecture | SAP

The infrastructure of a data warehouse can be on-premises, physically present in the company and managed by internal resources, or in the cloud, thus outsourced and characterized by greater scalability. The architecture of these data storage solutions typically consists of three levels:

- **Lower level or data level**: it consists of the server and the relational database system that collects, cleans, and transforms data from various sources according to the ETL (extract, transform, load) process. This is where data is stored and optimized, ensuring faster query times for frequently accessed data. Metadata is also created at this level.
- **Intermediate level or semantic level**: here is the OLAP (on-line analytical processing) server, which constitutes the analysis engine, used to access and analyse data through high-speed queries. OLAP is an approach to decision-making processes that allows cause-effect analysis, and it is characterized by a structure that enables navigation between information, using hierarchies and relationships to follow logical paths from general to specific according to efficiency principles.
- **Upper level**: the front-end interface where processed data are visually presented, allowing the end user to interact with the data, create dashboards and reports, and conduct ad hoc analyses. Often associated with a workbench or sandbox area for further data exploration, this structure relies on four components that, over the past decades, have determined the success of using structured data and SQL coding as the standard and more rapid way to interact with them:
    - ***Database***: typically in the cloud - following the advent of big data – it is the origin of the data contained in the data warehouse. Data can be stored in a central archive or in smaller ones, where only specific data from a particular business unit or department is allocated, speeding up information retrieval and obtaining more targeted insights thanks to greater granularity. Both in warehouses and marts, data access is regulated by security and governance criteria. Generally, it is worth noting that the organization of data in this type of infrastructure can typically follow two schemas:
        - Star schema, simpler and more frequent, because it is faster in query execution, where a table is joined to others of denormalized dimensions.

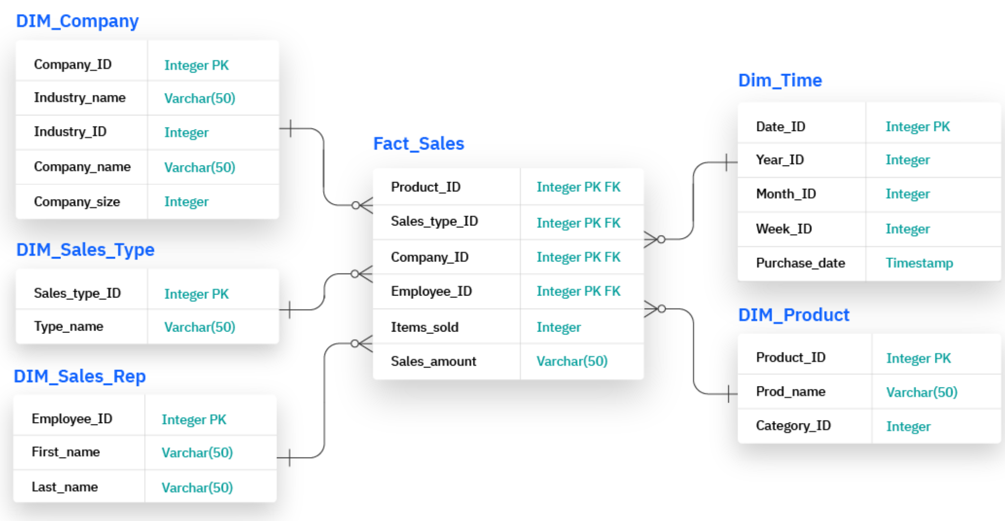



*Figure 2. Star schema. Source: IBM*

- Snowflake schema, slower in execution but less redundant, where a table is linked to others of normalized dimensions, which in turn are joined to "child" tables.

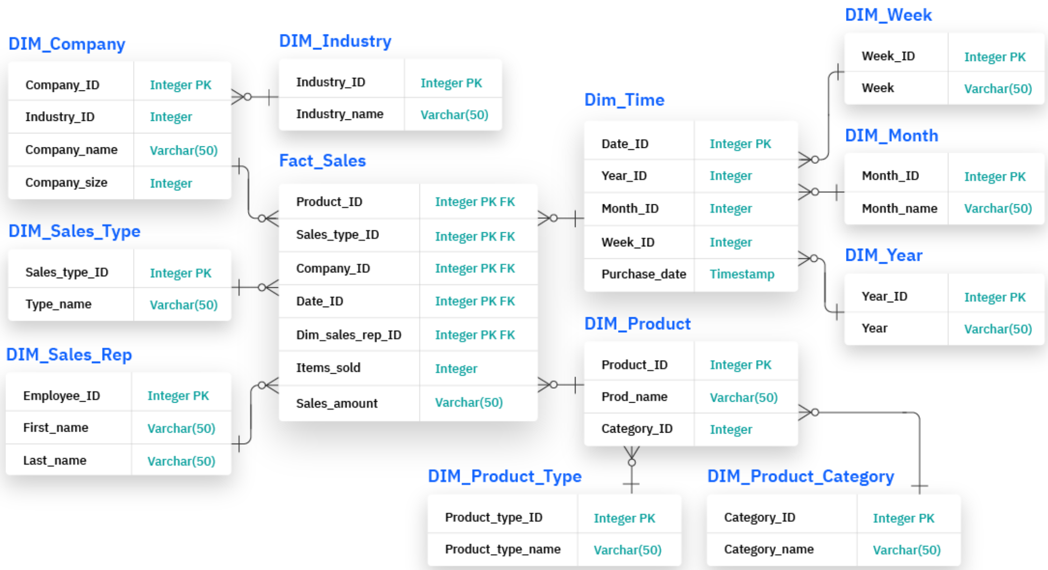



*Figure 3. Snowflake schema. Source: IBM*

- ***ETL process*** (extract, transform, load): converts data contained in the database into a usable format - suitable for analysis and queries - in the data warehouse. Data is extracted from their source systems and transformed (e.g., with groupings, translations, and derivations) to allow rapid analytical use of the information.
- ***Metadata***: they are data about the data stored in a system, describing them and making them searchable through characteristics such as origin, usage, value, etc. They can concern the context of the data or be more technical, explaining how to access them.
- ***Access tools***: enable front-end interaction with the data contained in the warehouse through querying, reporting, application development, data mining (based on SQL query processing).

Data warehouses allow the storage of current and historical data of a company, fuelling business intelligence through analysis processes and additional functionalities typically related to modelling. As a centralized repository of the entity's data, it can also be organized into sublevels, namely data marts, which contain only data of interest to a specific selected user group, thus speeding up query execution.

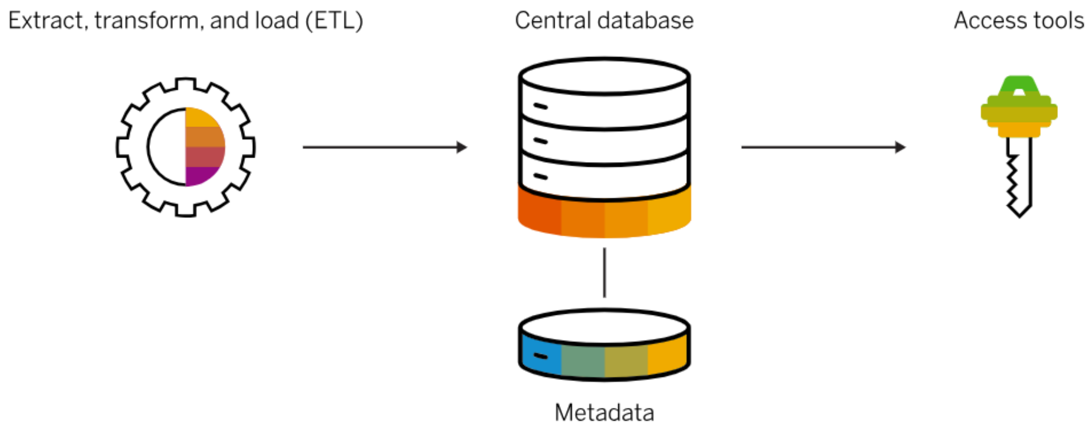



*Figure 4. Diagram showing the components of a data warehouse. Source: SAP*

Having **business intelligence** as **their main application area**, data warehouses can be particularly useful in the following contexts:

| | |
|---|---|
| Healthcare | these architectures offer a controlled environment to securely store and analyse patients' historical data to improve service quality and reduce the risk of errors. |
| Financial sector | the query speed enabled by data warehouses can be ideal for monitoring transactions and detecting potential fraud. |
| Marketing operations | the availability of processed and structured data stored in a data warehouse is extremely useful for improving the company's commercial performance |
| Logistics and supply chain | the ability to host specific data marts with faster controlled access makes the warehouse particularly useful for inventory management and supply chain optimization. |

The widespread adoption of this architecture in recent decades as well as nowadays, with data warehouse as-a-Service in cloud solutions, is due to some advantages recognized as strategic by companies:

- **Better quality and greater reliability of data**, because they are stored after cleaning, deduplication, and standardization operations, transforming them into a format compatible with subsequent analyses, to gain more complete and accurate information.
- **Long-term storage of historical data**, useful for identifying trends, learning from past issues, and making predictions to improve the business.
- **Smarter internal decision-making processes** made possible by data mining, visualization tools, AI and machine learning applications supported by the data warehouse, and - more generally - by faster queries thanks to a structure that facilitates the analysis of substantial amounts of data and the effective retrieval of requested information.

As structured data archives, despite the significant speed of analysis execution, data warehouses can often have **excessive costs associated with infrastructure maintenance and solution scalability** as the company grows and the volumes of data to be managed increase. It is also necessary to remember that **governance of the tool is essential** to ensure the profitable integration of data from various sources, with the awareness that maintaining accurate, complete, and up-to-date data can require significant integration efforts.

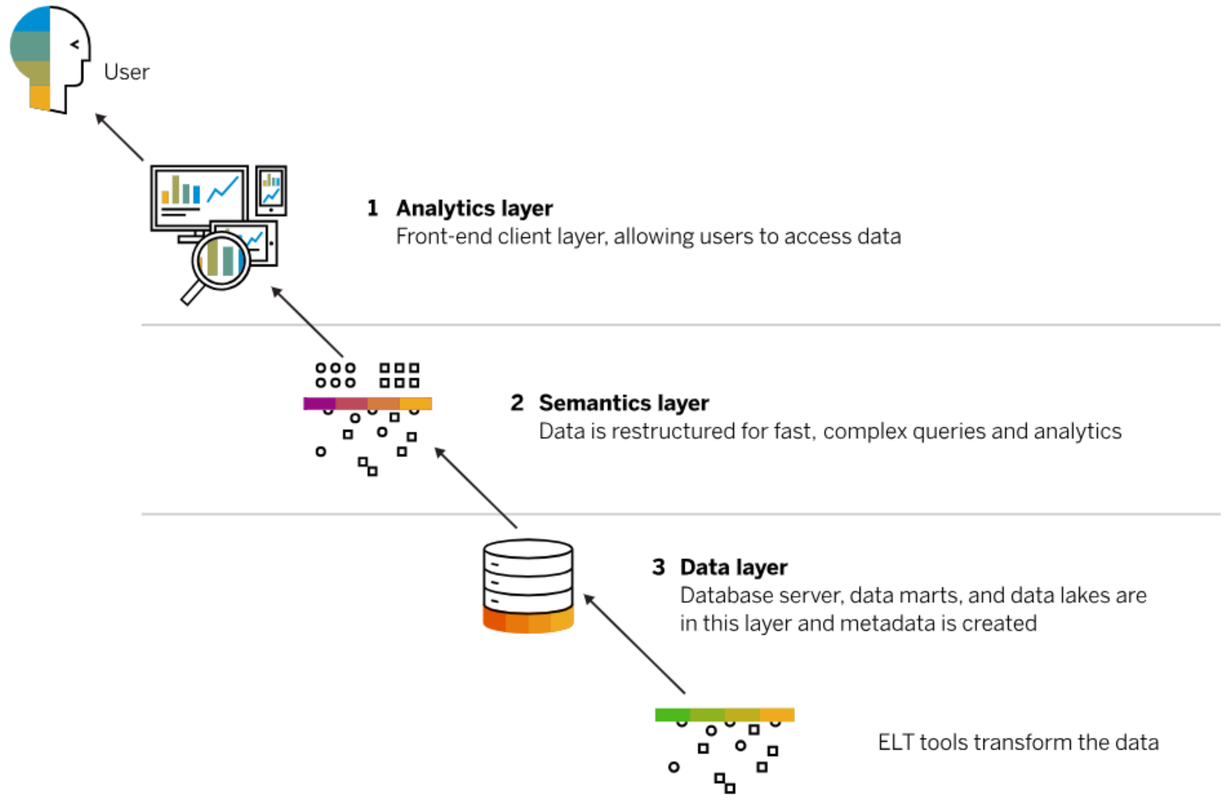



*Figure 5. Diagram of data warehouse architecture. A typical data warehouse includes the three separate layers above. Today, modern data warehouses combine OLTP and OLAP in a single system. Source: SAP*

## Data Lake

Data lakes[2] are **centralized repositories that allow the storage of large volumes of data in their native format and their use for a wide range of analysis needs**. What distinguishes this solution from data warehouses is that the data stored in this capacious and scalable architecture can be structured (like tabular data in a database) or unstructured (like images and audio files). **The absence of a predefined schema for storing data ensures optimal flexibility** for performing data discovery, data mining, data visualization, and machine learning projects: the data can be organized as needed depending on the type of analysis to be performed. Like data warehouses, the infrastructure of a data lake can be on-premises or in the cloud.

[2] Sources for this section:
What is a Data Lake? Data Lake vs. Warehouse | Microsoft Azure
What is a Data Lake? – Introduction to Data Lakes and Analytics - AWS
What Is a Data Lake? | IBM
What Is a Data Lake? | SAP

The main functionalities of this architecture concern data transfer, storage, and analysis:

- **Transfer**: data lakes allow the import of large amounts of data, even in real-time and from different sources, in their original format. This peculiarity saves the time that would otherwise be necessary to define structures, schemas, and required transformations.
- **Storage**: data lakes store relational and non-relational data, providing each element with an identifier and a set of accompanying metadata. Through indexing and cataloguing, it is always possible to understand what data is in the repository.
- **Analysis**: thanks to data lakes, various roles within organizations can perform analyses using sandboxes to develop proof of concepts without having to move the data to a separate system. It is possible to easily configure and reconfigure models, queries, and live apps and proceed with data analytics more flexibly.

The ability to store large volumes of semi-structured and unstructured data makes data lakes particularly advantageous for creating proof of concepts and training machine learning models. These architectures can be optimal, especially **when the business purpose for which the data is collected and stored is not defined in advance but can evolve over time**. The low storage costs and the preservation of data in its native format also make data lakes a valuable tool for data backup and disaster recovery.

**Data lakes and data warehouses can thus be complementary**: raw data stored in a lake could later be used to answer a business question and undergo an ETL process to make it usable within a warehouse architecture.

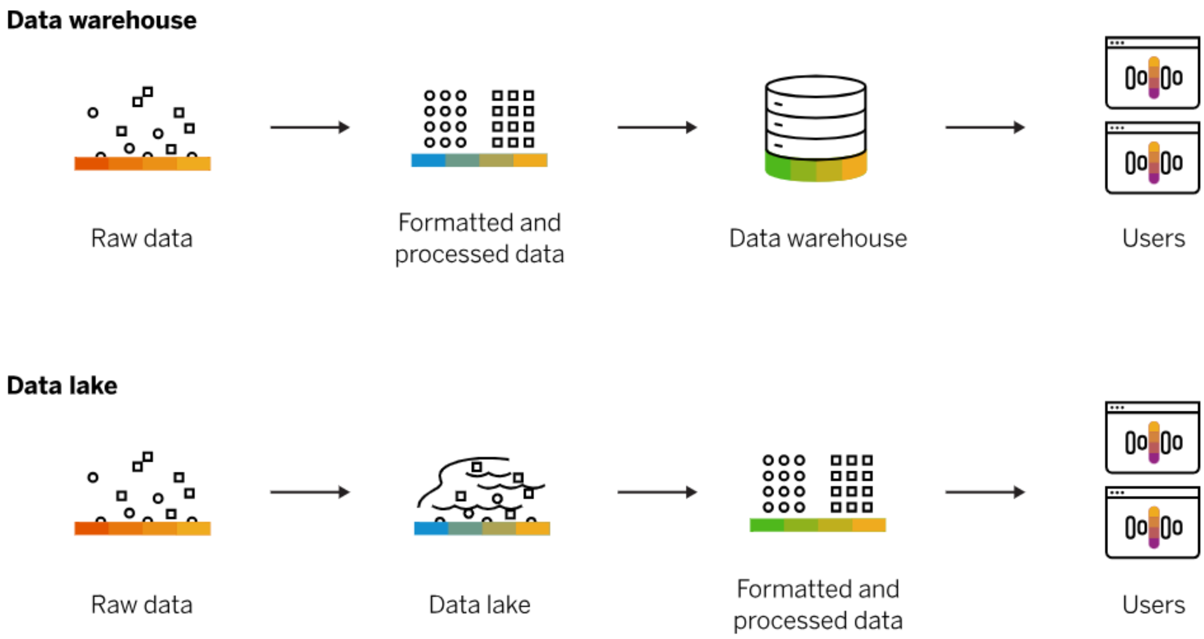



*Figure 6. Diagram of a data warehouse compared with a data lake. Source: SAP*

Given the characteristics listed above, there are numerous use cases for these architectures:

| | |
|---|---|
| Banking and financial sector | used to analyse transaction data and to efficiently manage portfolio risks by investment companies, which need up-to-date market data stored in real-time. |
| Healthcare sector | data lakes allow the management of big data and improve the quality of patient care and optimize hospital operations. |
| Energy and utilities sector | enable the storage of real-time data from energy networks and associated IoT sensors, allowing the study of consumption and more efficient resource management. |
| Retail and e-commerce | the storage of large volumes of several types of data enables to analyse customer purchasing habits, manage inventory, and improve marketing strategies. |
| Entertainment sector | streaming companies collect and process detailed information on customer behaviour and, thanks to data lakes, can improve the performance of recommendation algorithms and enhance the personalization of the user experience. |

The flexible architecture of data lakes can thus be particularly advantageous in terms of:

- **Information valorisation**: by storing large volumes of several types of data, queries can access all available information, regardless of the source that generated it, and apply a wide variety of tools to best valorise it.
- **Improving customer interaction**: data lakes reduce time-to-market thanks to always immediate and real-time access to information and by enabling interaction between data from different sources, helping to define a more effective business strategy and improve market positioning, also bringing value to PoCs created with stored data.
- **Reducing storage costs**: with a traditional system, it is necessary to foresee all the uses that will be made of the data in advance and therefore the ways in which it will need to be stored. With data lakes, problems related to changing business needs and analysis requirements that would require database restructuring and volume increase are circumvented.

At the same time, when opting for a data lake, it is necessary to be aware that **performance will typically be slower** than other data systems, lacking a rigid storage structure. Additionally, for data to be usable for further analysis and studies, internal mechanisms within the architecture must **ensure access controls, semantic consistency, and an effective cataloguing system**. Proper management requires solid governance of the tool, minimizing the risk of turning the data lake into a data swamp, a disorganized storage incapable of verifying the quality and consistency of the data within it, making it unusable.

## Data Lakehouse

Data lakehouses[3] have emerged in recent years as a new hybrid architecture capable of combining the advantages of data warehouses and data lakes: they **leverage data structures and integrate the metadata level characteristic of warehouses, combining it with the low-cost storage and flexibility of data lakes**. In this way, data lakehouses allow the storage of large volumes of data (structured, semi-structured, and unstructured) quickly and efficiently and enable working with various high-performance programming languages.

In the past, data warehouses and data lakes had to be implemented as isolated and distinct architectures, thus avoiding system overloads; the former stored structured data for business intelligence and reporting, while the latter used unstructured data primarily for machine learning operations. The separation between architectures became problematic when the data stored in each had to be processed together, increasing complexity and costs, and causing issues with updates, duplication, and data consistency. Data lakehouses are configured as a possible integrated solution, characterized by flexibility and scalability, capable of improving analysis performance and generating greater value from the data stored within them.

In summary:

- **Data warehouses are optimal for business intelligence reports due to their ability to execute queries on structured and semi-structured data stored within them**. For this reason, they are not the most suitable tool for supporting machine learning operations.
- **Data lakes** aim at responding to the exponential increase in data types and volumes. As the result of technological progress in data storage, they **are extremely useful for storing and processing data of any nature and in any format, and thus for machine learning tasks**, but not for creating business intelligence reports.
- **Data lakehouses** attempt to **synthesize the advantages of the other two architectures** with indexing protocols optimized for machine learning and scientific data analysis, low query latency, and high reliability in business intelligence performance.

A fundamental characteristic that fosters the preference for data lakehouses is the fulfilment of **ACID properties in transactions** occurring within the architecture, ensuring data integrity. The properties in the acronym are as follows:

- **Atomicity**: since transactions are often composed of multiple instructions, atomicity ensures that each is treated as a single operation, not allowing partial executions. If one of the instructions guiding the transaction fails to complete, it fails entirely, and the underlying database remains unchanged.
- **Consistency**: this characteristic ensures that, following a transaction, no contradictions occur between the stored data - all must be valid according to all the rules and constraints defined within the database at the beginning and end of a transaction, thus limiting the risk of corruption.
- **Isolation**: thanks to this property, transactions executed simultaneously appear to be serialized. The intermediate state of the transaction is invisible to other transactions running concurrently, so the result of simultaneous operations on the database corresponds to what would have been obtained by performing them (with inevitably longer times) one after the

[3] Sources for this section:
What is a data lakehouse? - Azure Databricks | Microsoft Learn
What is a Data Lakehouse? | IBM
What is a data lakehouse, and how does it work? | Google Cloud

other. This also ensures that the failure of one operation does not interfere with the execution of others.

- **Durability**: following a successfully completed transaction, the changes made cannot be lost, and the modifications to the data persist and are not undone, even in the event of a system failure, further ensuring data consistency.

The complex functionalities of a data lakehouse find application in all the use cases mentioned for data warehouses and data lakes, as well as many other sectors, ranging from manufacturing to telecommunications, which can thus generate value from data by integrating the specificities of the two previously described tools.

The architecture of data lakehouses is highly structured and typically develops on five levels:

- **Acquisition level**: all raw data is collected here from their various sources (internal and external from the entity) and transformed into a format that can be stored and analysed within the lakehouse.
- **Storage level**: it is an archive for structured, semi-structured, and unstructured datasets produced at the acquisition level. Being decoupled from computing resources, it is possible to store large volumes of data at a cost-effective rate: however, some cleaning, deduplication, and normalization operations are still performed at this level.
- **Metadata level**: it is a unified catalogue that provides details on all stored objects, allowing the use of management functions such as ACID transactions, file caching, and indexing to speed up query execution.
- **API level**: at this level of the lakehouse, APIs (application programming interfaces) are used to increase task processing and conduct more advanced analyses.
- **Data consumption level**: here, apps and analysis tools are hosted. They allow access to data and metadata and use them for business intelligence, data visualization, and machine learning purposes. Data is often highly aggregated, filtered, and optimized for query execution.

If the main disadvantage of data lakehouses is the **complexity of building such a structure from scratch** (also because it is **still not widespread**) through the integration of the logics that guide data warehouses and data lakes, the associated benefits are numerous and could convince companies to adopt them:

- **Higher quality and reliability of data** thanks to the fulfilment of ACID properties in transactions and the ability to reduce data redundancy by storing them in a single system accessible to the entire company, limiting inconsistencies that can occur when data is moved through pipelines in different systems.
- **Wide variety of use cases**, as it is possible to connect multiple tools directly to data lakehouses to support workflows of business intelligence, data visualization, machine learning, data science, etc., with a single repository.
- **Economic efficiency**, related to storing enormous volumes of data in low-cost storage space, in a similar way to what happens in data lakes.
- **High scalability**, since - unlike warehouses - these architectures decouple storage from processing, allowing multiple users to access the repository to perform simultaneous calculations on different processing nodes.

- **Better governance**, thanks to the ability to work on a single platform (instead of two separate ones), ensuring that acquired and loaded data meet system requirements and governance and security controls.

The following schema summarises the main features of the three data storage architectures previously described.

| | Data warehouse | Data lake | Data lakehouse |
|---|---|---|---|
| **Type** | Structured<br><br>Relational | Structured, semi-structured, unstructured<br><br>Relational, non-relational | Structured, semi-structured, unstructured<br><br>Relational, non-relational |
| **Schema** | Schema on write | Schema on read | Schema on read, schema on write |
| **Format** | Processed, vetted | Raw, unfiltered | Raw, unfiltered, processed, curated, delta format files |
| **Sources** | Application, business, transactional data, batch reporting | Big data, IoT, social media, streaming data | Big data, IoT, social media, streaming data, application, business, transactional data, batch reporting |
| **Scalability** | Difficult and expensive to scale | Easy to scale at a low cost | Easy to scale at a low cost |
| **Users** | Data warehouse professionals, business analysts | Data scientists, data engineers | Business analysts, data engineers, data scientist |
| **Use cases** | Core reporting, BI | Machine learning, predictive analytics, real-time analytics | Core reporting, BI, machine learning, predictive analytics |

# Business models for data valorisation

After examining the characteristics of the main tools used for data storage and explaining which architectures may be more suitable for the growth goals that each entity, in its specific sector, may want to pursue, in the following pages a description of two solutions based on economic data valorisation will be provided. Data marketplaces and open data portals are platforms that collect data generated from various sources and make it available to private and public actors, allowing them to improve their services and performance, thus enabling the development of data-driven business models. The principles underlying these two solutions and the different types of data collected and hosted in these tools will be also fundamental for the creation of data spaces.

## Data Marketplace

Data marketplaces[4] are **online platforms** - typically cloud services - **where providers and consumers can list, purchase, and exchange data from various sources in a secure and controlled environment** where data quality and integrity are guaranteed. As anticipated, in the era of big data, most companies (especially the big ones) have realized that the data they generate is not just a byproduct of their operations but represents a valuable asset to be valorised for competitive advantage: they find in marketplaces a tool that allows them, on one hand, to **efficiently promote and sell their data** and, on the other hand, to **explore and purchase datasets from others**, all through easy-to-use self-service interfaces and search and filter functions.

Data marketplace platforms are an optimal solution when companies want to monetise their data: they centralise transactions and eliminate the complexities associated with searching for datasets and negotiating purchases, thus facilitating exchanges with a greater number of actors, ensuring security, and enabling increasingly data-driven decision-making processes thanks to immediate access to new information. For this reason, there are numerous types of data marketplaces, depending on the source of the data, the level of security provided, or specific goals guiding the exchange:

- **Public data marketplaces**: the data contained within these spaces are generated by public actors and are subject to service level agreements and possible usage restrictions based on local governance.
- **Private or personal data marketplaces**: these platforms remunerate individuals for sharing their data and thus allow consumers to have greater control and privacy over their personal information.
- **Multilevel data marketplaces**: in this architecture, raw, processed, or aggregated data can be included, and the various levels correspond to greater complexity to ensure exchange security.
- **B2B data marketplaces**: typically used by organizations to support their market growth.
- **IoT sensor data marketplaces**: these platforms are built on an architecture capable of integrating information flows from sensors, particularly useful for companies and entities interested in monitoring specific phenomena over time.

---

[4] Sources for this section:
Data marketplace | Interoperable Europe Portal
Data marketplace - Cloud Adoption Framework | Microsoft Learn
What is a Data Marketplace? | IBM
What Is a Data Marketplace? An Ultimate Guide for 2025

Although most marketplaces are externally oriented, there are also internal data marketplaces within organizations and companies that generate and use large amounts of data. In this case, they often use non-proprietary solutions that integrate with multiple technology stacks and are designed as tools that ensure all authorized users can quickly and effectively select the data they need to carry out the company's various projects: sharing is thus simplified, and data discovery operations are facilitated.

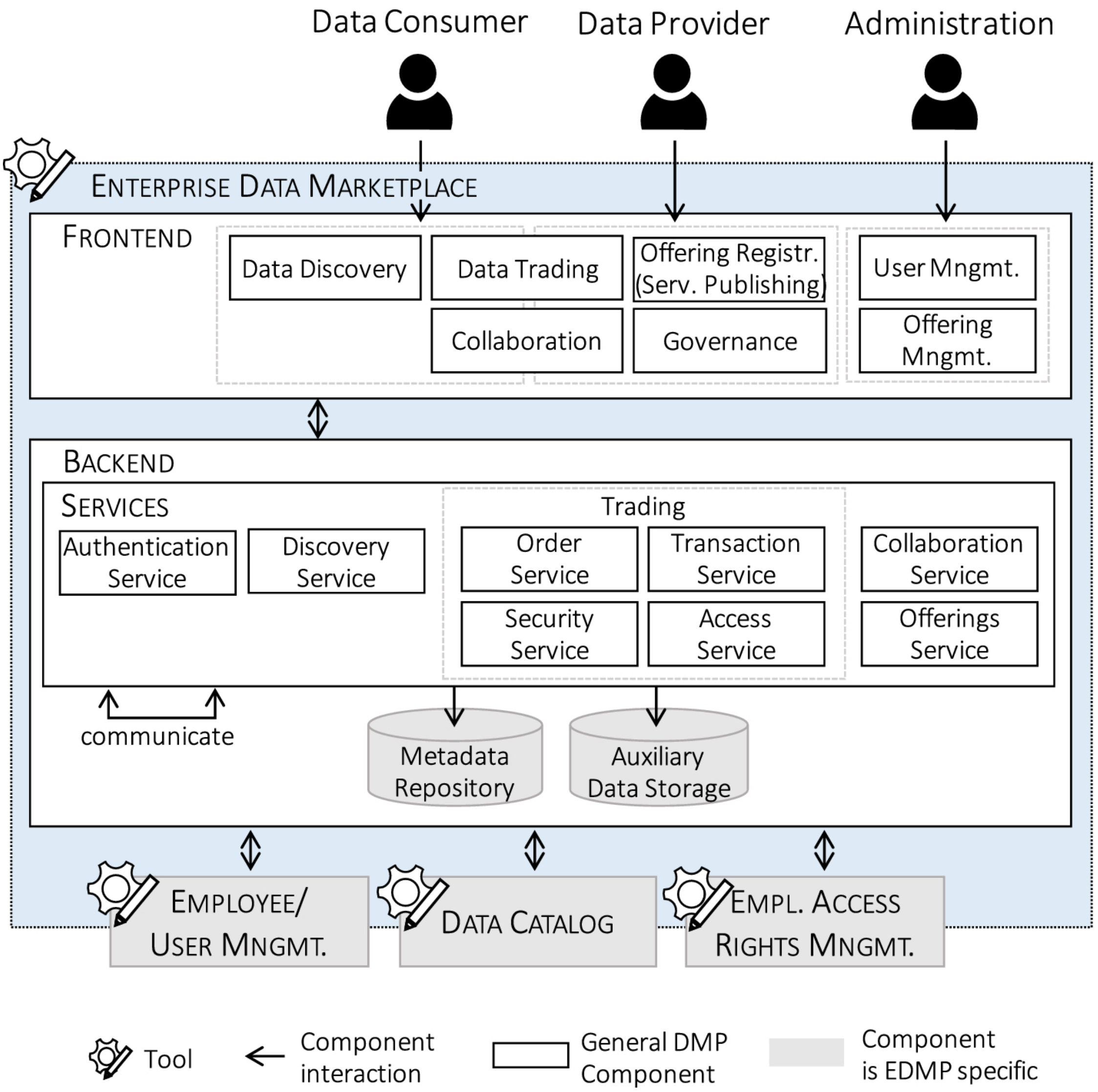



*Figure 7. Enterprise data marketplace architecture featuring a component overview. Source: Eichler, R., Gröger, C., Hoos, E. et al. (2023). Introducing the enterprise data marketplace: a platform for democratizing company data. Springer Journal of Big Data*

Regardless of the type, all data marketplaces are characterized by an ecosystem populated by different stakeholders:

- **Data providers**: companies, organizations, entities, and individuals who commercialize their data assets on the platform, setting the price, providing all necessary details about their data, and ensuring they are accurate and compliant with the marketplace's standards and policies.
- **Data consumers**: companies, organizations, entities, and individuals who seek datasets for various purposes (analysis, business intelligence, machine learning, etc.), relying on the platform's robust security measures and the guarantee of data integrity hosted there.
- **Platform operators**: those who own the marketplace and handle developing, maintaining, and managing the platform at the infrastructure and performance levels, integrating

necessary security measures, and ensuring data integrity. Platform operators support interactions between providers and consumers and often integrate into the architecture tools for searching and reviewing data through working on metadata.

- **Governance authorities**: actors who set up regulatory policies, quality standards for commercialized data, and compliance requirements with current regulations and privacy standards. Authorities check commercial interactions and intervene to resolve any compliance issues.

The architecture of data marketplaces is designed to address the various forms of interaction between the aforementioned stakeholders. Consequently, the following functionalities are included:

- **Data control**, in terms of quality, source, and format, to ensure they are fully accessible and usable by potential buyers.
- **Data discovery**, thanks to search tools based primarily on metadata catalogues that allow users to effectively identify what they need.
- **Application Programming Interface (API)**, which facilitates interaction between data providers and consumers, ensuring that only authorized users can access the data, in a secure and reliable way.
- **Monetization** made possible through the integration of transaction tools with which providers can license and sell their products.

Each of these functionalities corresponds to a fundamental requirement for the establishment and success of a data marketplace:

- ***Data quality and integrity***: implementing controls and quality assurance processes ensures that the marketplace hosts data standardized in format and structured according to a schema that allows consumer use. Data should be accurate, consistent, complete, durable, from reliable sources, and - where possible - anonymized.
- ***User accessibility***: the user interface must be intuitive and accessible, allowing searches of data and metadata catalogues through keywords and natural language. The ability to implement quality control and data review systems on the platform helps strengthen the data marketplace's reputation and increase provider and consumer loyalty.
- ***Security and privacy***: compliance with security protocols and national and international privacy laws prevents unauthorized data access and data breaches. Data marketplace governance also involves controlling how data is collected and whether proper consent for use has been obtained from the owners.
- ***Transaction management***: security is a concern not only for data but also for transactions and payment systems, with clear licensing agreements defining rights and limitations for both providers and consumers.

Data marketplaces act as centralized hubs where public and private entities can economically valorise their data. The numerous advantages associated with the utilization of these platforms, as outlined in the preceding analysis, can be summarised as follows:

- **Marketplaces simplify the process of acquiring data**, as they allow entities to quickly find and procure the data they need, saving time. Furthermore, they promote competition even for smaller entities, which would otherwise have to invest many resources (economic and otherwise) to obtain certain datasets, or risk not having access to some information sources.
- **More democratic access to data made possible by marketplaces supports companies' decision-making**, allowing them to define more informed and data-driven development

strategies, thanks to the availability of a wider range of data from other companies, different geographical regions, and various levels of granularity.

- Finally, **marketplaces act as innovation accelerators**, allowing companies to work on external data they did not possess and integrate it with internal data to develop diversified solutions and new approaches, also stimulating collaboration among actors present on the platform.

## Open Data Portal

Open data portals[5] are **platforms that facilitate the identification, access, and reuse of open data**, typically adopted by public administrations at local, national, and international levels to disseminate public sector information and promote its use by other actors, both private and public.

Initially developed to facilitate the organisation and dissemination of large volumes of data from public entities, open data platforms have also spread among companies, which make select information available to developers for the creation of new software applications. The web is the main platform where data can be uploaded and found; however, specific open data portals facilitate the data publication process and ensure consistency for consumers in terms of access, download, analysis, visualization, and data use (especially thanks to APIs and metadata-based cataloguing systems).

It is important to remember that data can be designated as open if it is accessible, utilizable - sometimes with the obligation to cite the source - and shareable. Therefore, it must be available in a common and machine-readable format, accompanied by a license that permits users to transform it, combine it with other information, and re-share it with others, adhering to the same license as the original open data, even for commercial purposes. **Interoperability**, the ability of different systems and organizations to work together, **is essential to realize the main advantage of open data**, as it allows the combination of different databases and thus the development of new products or the improvement of existing ones. Data containing information about individuals or those related to national security cannot be published as open data; however, all other categories of data can be open data, provided they meet the following characteristics:

- **Completeness**: they must include all components that allow them to be exported, used online and offline, integrated, and aggregated with other resources (for this reason, they must be presented in a sufficiently disaggregated manner), and disseminated on the web.
- **Accessibility and searchability**: they must be transmissible and interchangeable among all users on the web directly through Internet protocols, without any contract subscription, payment, registration, or official request. Open data must also be fully accessible to machines (machine-readable), i.e., automatically processable by computers, and easily searchable through specific search tools, such as databases, catalogues, and search engines.
- **Non-proprietary**: consumers must be able to use and process data through non-proprietary programs, applications, and interfaces, without any limitation on use and reuse - copyright and intellectual property rights cannot be implied - including for commercial purposes and for creating new resources, applications, and services. Open data should be made available for free or at a price not exceeding a reasonable reproduction cost.

[5] Sources for this section:
Open data portals | Shaping Europe's digital future
Getting to grips with platforms
The Open Data Handbook

- **Timeliness and permanence**: data should be made public as quickly as necessary, in order to facilitate the extraction of value from it. Moreover, data should be kept consistent and updated, whilst always respecting the other characteristics described above.

Platforms that collect and make open data available can follow two different approaches:

| | |
|---|---|
| They can act as **data catalogues** and thus list the datasets available on the web, often classified by sector or topic; these platforms allow users to know what open data exists, what it concerns, where it is located, and how it can be obtained, acting as a link between users and data. | They can be more properly **open data management platforms** and thus primarily target those who want to publish data and keep it consistent and up to date over time, also providing support in structuring and formatting data. Users, in turn, can search and process the information they need directly in these spaces without having to download any dataset. |

In both cases, these portals are oriented towards achieving some primary goals:

- **Finding data**: platforms are designed to allow users to quickly find and reuse the data they are interested in, with a preference for the most relevant resources. Thanks to data feeds provided by the platforms and indexable across the web, search engines can more easily find the required open data.
- **Offering a consistent user experience**: the configuration of the platforms is designed to facilitate the user experience of searching. For this purpose, many public administrations have also adopted a standard convention for the nomenclature of their portal's web address.
- **Accessing data consistently**: the establishment of open standards by the web community has resulted in the presence of consistent APIs through which finding, accessing, and using data (and metadata). This development is compounded by the adoption of similar user interfaces across diverse platforms.

In accordance with the European data strategy, in 2021 the European Union launched **data.europa.eu** as a single access point for open data published by European institutions, national portals of member states, and international organizations active in the European context. The goal is to stimulate the cross-border dissemination of reusable information within the Union, improving the discoverability and accessibility of information and thus supporting the development of new applications and products based on a greater and better quantity of data and responding to common needs. The core of the portal is represented by the **common format metadata catalogue** through which it is possible to **access over 1,600,000 available datasets, mostly freely usable for commercial and non-commercial purposes** - specific conditions for reuse related to intellectual property and/or privacy protection are also indicated. data.europa.eu also publishes studies and reports, including an annual assessment of the open-data maturity of each contributing state, and integrates a data.europa academy section that promotes open data literacy through free courses, videos, and self-training tools.

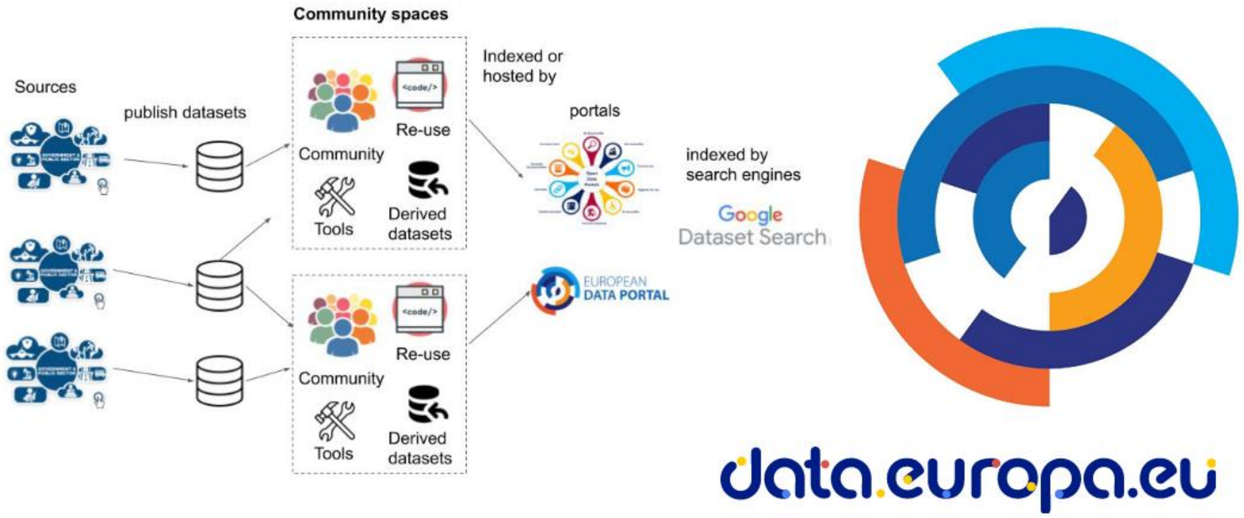



*Figure 8. data.europa.eu ecosystem. Source: dati.gov.it*

In 2023, as part of this European initiative, the study *New Business Models for Data-Driven Services* tried to explain how information generated by the public sector and released as open data could be reused by private for-profit and non-profit actors, generating value for society and the economy, precisely through **public-private collaboration**. To this end, it was highlighted how the size of the **open data market**, which for the 27 Union countries **in 2019** was **equivalent to 184.45 billion euros**, is expected to **grow by 2025, ranging from 199.51 billion in a baseline scenario to 334.2 billion in a more optimistic scenario**. 12% of value creation in the data economy directly derives from open data, and 45% is indirectly impacted (e.g., in synergy with other data sources): both the public and private sectors should therefore give this mode of data valorisation their utmost attention, particularly in view of the fact that the areas expected to see the greatest impact from open data range from education to transportation, commerce to finance, the energy market to healthcare. They are all sectors in which collaboration and synergy between public and private are desirable and the main benefits associated with leveraging open data (the creation of new services, the improvement of decision-making capacity, and increased efficiency through business operation optimization) can be profitably achievable, also using technologies such as artificial intelligence and blockchain.

The study of the **open data value chain** allows for a greater appreciation of the advantages associated with the spread of open data platforms. The data flowing into these platforms is generated for internal service provision purposes or in compliance with a legal obligation, and their publication - not being a primary objective - constitutes a strategic element aimed at generating new economies. Since this data often requires a continuous flow of generation and preservation, it is necessary to consider the costs associated with their maintenance and modernization, often borne by actors other than those who generate value through the reuse of such information. As shown in the following diagram[6], the main activities characterizing the open data value chain are **generation, dissemination, retrieval, storage, classification, exposure, reuse, and consumption**, which produce - as results - first raw data, then refined data (as made available on platforms), and finally products and services oriented towards a specific purpose. The involvement of governmental and private actors, for-profit and non-profit, supports the need for a cooperative approach that facilitates the creation of European ecosystems capable of sustaining the economic growth described above.

[6] Charalabidis, Y., Zuiderwijk-van Eijk, A., Alexopoulos, C., Janssen, M., Lampoltshammer, T., & Ferro, E. (2018). *The World of Open Data: Concepts, Methods, Tools and Experiences.* (Public Administration and Information Technology; Vol. 28). Springer, p. 119

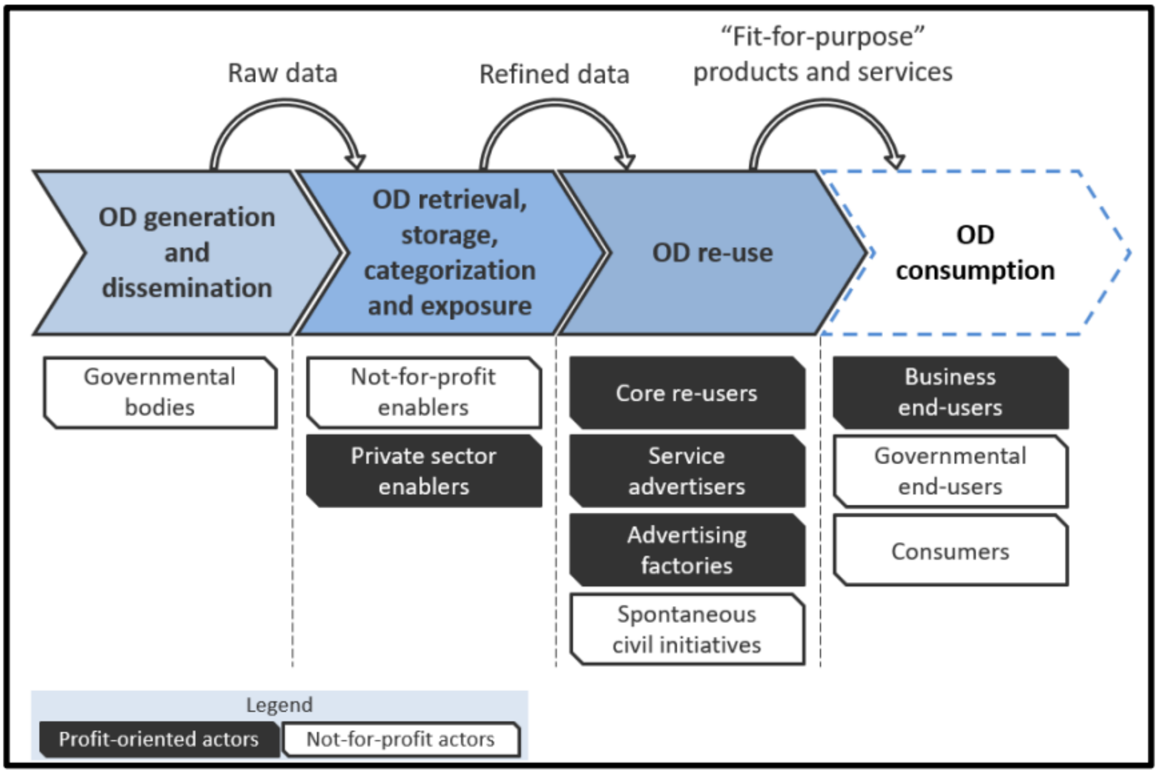



*Figure 9. Open data value chain. Source: Charalabidis, Y., Zuiderwijk-van Eijk, A., Alexopoulos, C., Janssen, M., Lampoltshammer, T., & Ferro, E. (2018). The World of Open Data: Concepts, Methods, Tools and Experiences. (Public Administration and Information Technology; Vol. 28). Springer*

To this end, it is becoming increasingly clear that there is a need to create technical, legal, and procedural preconditions, as well as to identify adequate business models that can ensure the long-term financial sustainability of these activities. The core of the business model obviously resides in the data released in accordance with the open paradigm and thus made accessible to all, without technical, legal, or price constraints. This raw resource is then subjected to processing, resulting in a distinct and more specific asset for the company, enabling data reuse in one of its products or services.

As suggested by the aforementioned study, the definition of the business model must first clarify whether the goal is "better business through open data," and thus whether open data is used instrumentally to improve the quality of a market good already present in a specific sector, or whether the ultimate goal is "open data pure plays," meaning whether such data will allow structuring and delivering services that otherwise could not exist, even spanning different areas. To resolve this first dichotomy and subsequently define the value proposition, it will be necessary to answer three questions:

- What types of data processing (data aggregation, classification, geo-referencing, visualization, AI algorithm training) will be used to obtain relevant information?
- What will be the role of open data in the entity's value proposition? Will they be the main market asset, an essential element of the products or services sold, or an accessory component?
- How do you intend to capture part of the created value? Will you opt for a premium, free, or freemium solution (often considered the most profitable)?

To the private sector, open data portals represent a tool for value creation from this type of data, according to a business model based on the suggestions presented above. However, they can also operate at a different level and simplify the search, integration, and processing of that data, by removing barriers that hinder reuse and affect providers and consumers. Platforms are indeed required to interface with a number of constraints, including institutional (e.g. right trade-off between public values, risk-averse culture), legislative (privacy and security), technical (e.g. absence of metadata, machine unreadability), data management complexity (difficulty of access, lack of clarity on the license, absence of quality information), and usage and participation constraints (absence of incentives for users). Reducing and resizing these obstacles through public interventions, also aimed at supporting the spread of an approach based on the principle of "open-by-design," is essential to foster the growth of the open data economy, as it will allow private entities to dedicate fewer resources to the data acquisition process in favour of greater investment in differentiating the value proposition compared to competitors. In this regard, open data portals can play a strategic role if primarily oriented towards **improving the usability of economically easier-to-reuse data**, such as geographical, meteorological, environmental, economic, and sociodemographic information: the most successful initiatives are those where the **goals of data providers are aligned with those of data re-users**, and it is therefore important that platforms also serve as spaces where these shared objectives can be identified.

Ultimately, the spread of the open data paradigm – also essential for the development of data spaces - cannot be separated from the creation of technological infrastructures. Portals and platforms promote a shift in the perception of data, transforming it from a cost to a strategic opportunity. This change entails the transition from the mere compliance with legal obligations to proactive provision of data to meet operational needs, both in the public and private sectors. Indeed:

| |
|---|
| The publication of open data creates value even for small businesses, offering opportunities for innovation and attracts new customers and new markets. |
| The combination of data from various sources creates greater value, capable of more effectively addressing complex challenges. |
| Services built on open data are valuable for a wide range of organizations and individuals, bringing benefits to various categories of stakeholders. |

For these reasons, in the European context, the impact assessment of initiatives related to open data develops along three axes: **political** (measuring the impact on the efficiency and effectiveness of political action and transparency), **economic** (with an indicator based on the assessment of the market value of open data and the improvement in the provision of services connected to them), and **social** (related to the impact on environmental sustainability and the inclusion of marginalized groups).

# How to… data valorisation: success stories form the Alpine regions

Before moving on to an in-depth examination of the data space approach as another data valorisation strategy, two examples of experiences based on the architectures described so far will be presented. The selected case studies insist on the Alpine regions and concern business models that are currently in vogue: Digital Twin and Mobility as a Service.

A **Digital Twin**[7] is an **integrated data-driven virtual representation of real-world entities and processes, with synchronized interaction at a specified frequency and fidelity**. Digital Twins use real-time and historical data to represent the past and present, and simulate predicted futures, and transform business by accelerating and automating holistic understanding, continuous improvement, decision-making, and interventions through effective action.

On the other hand, **Mobility as a Service**[8] (MaaS) integrates **various forms of transport and transport-related services into a single, comprehensive, and on-demand mobility service**. MaaS offers end-users the added value of accessing a diverse menu of transport options (including public transport, active modes such as walking and cycling, ride/ car/bike-sharing, taxi, and car rental or lease, or a combination thereof) through a single application and a single payment channel (instead of multiple ticketing and payment operations). A successful MaaS service also brings new business models and ways to organise and operate the various transport options, with advantages for transport operators including access to improved user and demand information and new opportunities to serve an unmet need.

The aim of these two examples is clearly not to provide an exhaustive description of the infinite possibilities available to private and public entities seeking to create value from the data they produce or collect. Instead, the objective is to raise awareness of the centrality of data by highlighting two good practices that are generating positive impacts at the local and regional level. These examples thus serve as a source of inspiration for other innovative solutions.

## The Digital Twin of Alps (DTA)

### *Context and description of the project*

The Digital Twin of Alps is an initiative developed within the framework of environmental and territorial digitalisation. The consortium behind this digital twin includes several key actors across Alpine space, such as Sentinel Hub GmbH in Austria, Sinergies d.o.o. in Slovenia, EOST/A2S Université de Strasbourg in France, Terranum Sàrl in Switzerland, MobyGIS and CNR-IRPI in Italy.

The project consists of a platform focused on two main themes: Water Resource Management and Disaster Risk Management, which represent the main environmental menaces and the most impactful dangers affecting the population of the Alps.

As part of ESA's Regional Initiative 3, the primary goal of the Digital Twin of Alps (DTA) is to **provide a roadmap for the implementation of a future Digital Twin Earth (DTE)**, which will be based on the integration of Earth observation data, interconnected numerical simulations and Artificial Intelligence.

The DTA enables the **monitoring of natural phenomena** such as landslides, floods, and climate changes, providing predictive tools useful for the land planning and management. Through the

[7] Definition of a Digital Twin
[8] What is MaaS? – MAAS-Alliance

integration of advanced simulation models and the use of the AI, it is possible to **analyse large volumes of data and generate future scenarios to support evidence-based decision-making processes**.

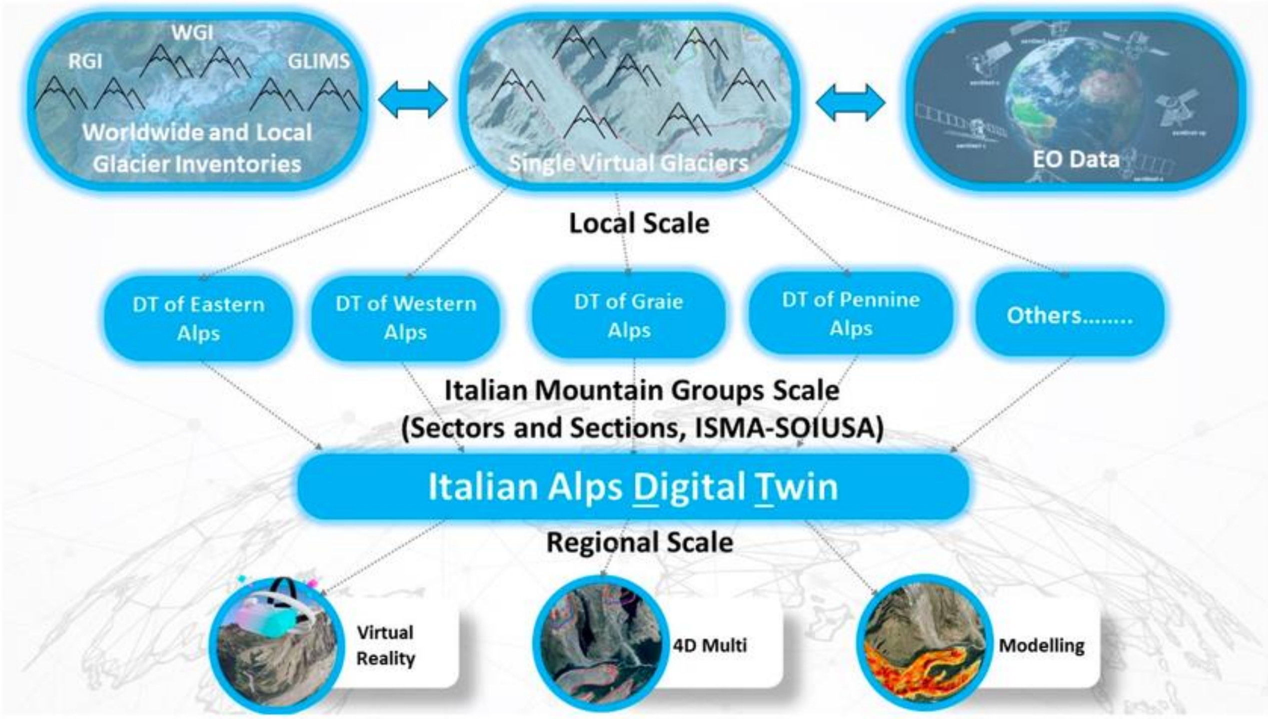



*Figure 10. Digital Twin of Alps theoretical framework: a compartmental system where local models interact to generate a global response at the Alpine level, providing different approaches to output depending on user's profile. Source: Fissore V., Bovio L., Perotti L., Boccardo P. (2023). Towards a Digital Twin Prototype of Alpine Glaciers: Proposal for a Possible Theoretical Framework, Remote SensingWater Resource Management in the Alps*

Most Alpine ecosystems are heavily reliant on sufficient water resources. Also for this reason, the Alpine region is known as the "Water Tower of Europe", due to its significant contribution to the formation of many major European rivers. This underscores the necessity for a decision-support tool.

The DTA focuses on providing ready-to-analyse datasets that benefit from a combination of inputs, including data from the Earth Observation (EO) and other sources. Furthermore, the DTA offers easy access to multiscale representation of complex hydrological processes and their interactions.

### Water Resource Management

The digital twin specifically focuses on three main challenges:

1. Snow cover: this process collects and processes meteorological reanalysis input data from ERA5, provided by European Centre for Medium-Range Weather Forecasts (CMWF) and from the Weather Research & Forecasting (WRF) model, with the objective of addressing data gaps. Algorithms are used to combine different information to create Machine Learning models. The input data, along with the maps, are then fed into the distributed physical model GEOtop, which undertakes the reconstruction of the snow cover evolution.
2. Water discharges and water heights: in this case, a geomorphological data extraction is performed, and input information is then collected and processed to derive variables such as precipitation and average temperature. Along with a geometric component, these water-related variables are then inserted into the semi-distributed NewAge model, which is used to simulate the real situation within the reference catchment area.
3. Soil moisture and drought monitoring: in this case, input data include satellite data on soil moisture, evapotranspiration, and precipitations, which are compared with the results of

hydrological modelling chain. The outputs consist of monthly maps of standardized anomalies, which allow the monitoring of water scarcity conditions from different components of hydrological cycle.

### Disaster Risk Management in the Alps

Due to a range of natural risks caused by geographical and climatic features of the Alpine region, such as avalanches, landslides, and floods, in the last decades a series of mitigation measures has been implemented. The DTA aims to provide a decision-support system for all stakeholders involved in monitoring and mitigating environmental risks and their impacts in the Alps, by combining data from *in situ* and remote observations, citizen surveys, and models for quantitative risk assessments.

The main challenges addressed by this digital twin in the context of environmental risk monitoring are:

- Flooding: maps identifying flood-prone areas are developed and stored in the system. Predictions of flooding events can be supported in real time thanks to the discharge hydrograph predicted by the hydrological model based on weather forecasts.
- Shallow landslides: in this case, the service employs a combination of models. The Shallow Landslide Failure Forecasting (SLFF) scheme aims to assess the probability of landslides, combining static spatial maps and water-related variables. The Shallow Landslide Flow (SLF) model, on the other hand, aims to evaluate the areas of landslide sources detected by the SLFF. This assessment is based on Flow-R, a software for modelling geological hazards.
- Large, deep slope failures, mostly deep-seated landslides, and rockslides: the deep-seated landslide processing chain provides displacement and velocity data, relying on GDM-OPT-SLIDE, a service enabling on-demand processing of optical Sentinel-2 imagery through image correlation algorithms. The GDM-OPT-SLIDE is available on the Geohazards Exploitation Platform (GEP) and provides crucial information to stakeholders responsible for landslide early-warning systems and risk management.
- Terrain motion: terrain motion products are derived from SNAPPING (Surface motioN mAPPING), an on-demand service for Sentinel-1 IW SLC (Interferometric Wide Swath Single Look Complex) based on SNAP and StaMPS, two integrated chains, which are available through the Geohazards Exploitation Platform (GEP). This service is pivotal for stakeholders involved in disaster risk management.
- Glacier motion: the service relies on Sentinel-2 images and the GDM-OPT-ICE web service. Glacier motion products available on the DTA demonstrator are based on image correlation algorithms applied to Sentinel-2 imagery, using the GDM-OPT-ICE web service accessible on the Geohazards Exploitation Platform (GEP).

### *Data Sharing Model*

The Digital Twin of Alps is part of a broader context in which data sharing between public, private and academic entities play a crucial role in managing environmental challenges. These data come from a wide variety of sources, including satellites, ground observation, IoT sensors, geospatial databases, and climate models, then collected and integrated into the digital infrastructure that makes them accessible to various stakeholders. They thus can be either open, such as maps, used to promote collaboration and foster innovation, or closed, when they involve information subject to regulations or proprietary data from specific entities.

Data interoperability is a key aspect of the project, ensuring that information can be shared across different systems and used effectively by the various stakeholders involved, including public administrations, universities, research centres and private companies.

Another crucial aspect of the DTA is the data sharing platform, which allows users to access detailed representations of the Alpine territory and interact with the data in an intuitive way. This platform fosters the collaboration among different entities and enables the development of innovative tools for the sustainable land management. For example, the data collected can be used by researchers to study the impact of climate change, by public entities to plan natural risks mitigation strategies, and by companies to develop technological solutions in the environmental and infrastructure sector.

*Actors Involved*

Data Contributors:

- **Earth observation satellites**: provide satellite imagery and geospatial data.
- **Weather agencies**: provide meteorological reanalysis and weather forecasts.
- **Ground observations networks**: contribute with data from sensors and field measurements.
- **Research institutions**: contribute with scientific data, models, and analyses.
- **Private companies**: provide modelling solutions and proprietary data.
- **Citizen surveys**: contribute with data on local environmental observations and perceptions.

Users:

- **Public Administrations**: use the platform for land planning and disaster risk management.
- **Researchers**: use data, for example, for environmental studies, climate change research and hydrological modelling.
- **Private companies**: use data for developing innovative and resilient technological solutions.

*Examples of Shared Data*

As previously shown, the data shared within the Digital Twin of Alps ecosystem can be classified in the following main categories:

- **Satellite Data**, which provide geospatial data for various environmental analyses.
- **Hydrological Data**, such as precipitation, snow cover and water discharge data, often used to forecast water resource availability.
- **Environmental risk Data**, e.g. maps of environmental risks.

## MaaS ToMove

*Context and description of the project*

MaaS ToMove is an innovative project by the City of Turin, coordinated by the publicly owned company 5T, providing residents of Turin and the metropolitan area with exclusive benefits to experience a new way of moving according to the Mobility as a Service (MaaS) paradigm. The primary goal of this initiative is to introduce the opportunities offered by innovative technologies, and promote integrated, conscious, multimodal, and sustainable mobility choices.

Through the MaaS service, participants can access a wide range of digital mobility services to meet any travel need through one of the available “Super Apps”, which offer an integrated travel experience and feature specific characteristics, both in terms of available services and user experience. Each participant can select the Super App that best suits their needs (it can also be an app already present on their smartphone, such as a taxi service).

Thanks to 5T's expertise in Mobility as a Service, the City of Turin has collaborated with all sector operators active in the area to support the ongoing digital transformation of mobility. Starting with its synergy with GTT (Gruppo Torinese Trasporti), whose public transport network (buses, trams, and metro) serves as the backbone of MaaS, Turin has facilitated the integration of the "MaaS ToMove" Super Apps with services from numerous mobility providers, such as TaxiTorino, RideMovi, Dott, Voi, Cooltra, and Flibco.

The MaaS ToMove project is an integral part of ToMove, the City of Turin's initiative focused on developing new scenarios for smart and sustainable urban mobility. This pilot project will guide Turin toward an increasingly integrated, efficient, and sustainable mobility future over the next two years. The entire initiative is funded through the Complementary Fund to the National Recovery and Resilience Plan (PNRR) as part of the national "MaaS4Italy" program, promoted in collaboration between the Department for Digital Transformation (DTD) of the Presidency of the Council of Ministers and the Ministry of Infrastructure and Transport (MIT).

In a nutshell, MaaS ToMove is not a standalone platform, but an **experimental initiative** aimed at integrating and incentivizing multimodal mobility solutions. To date, there are more than 6600 registered users who have taken more than 30,000 MaaS trips. **The project aspires to become a permanent service,** with **governance led by the city administration**. Currently, **market operators act as intermediaries**, managing incentives for citizens and users. Since the aim is to define a sustainable business plan for the initiative, the City of Turin and 5T are currently evaluating different alternatives, and it is likely that the development model of MaaS ToMove will be the result of a combination of targeted solutions for different possible types of users (e.g. subscription, pay-per-use, agreements with companies, etc.).

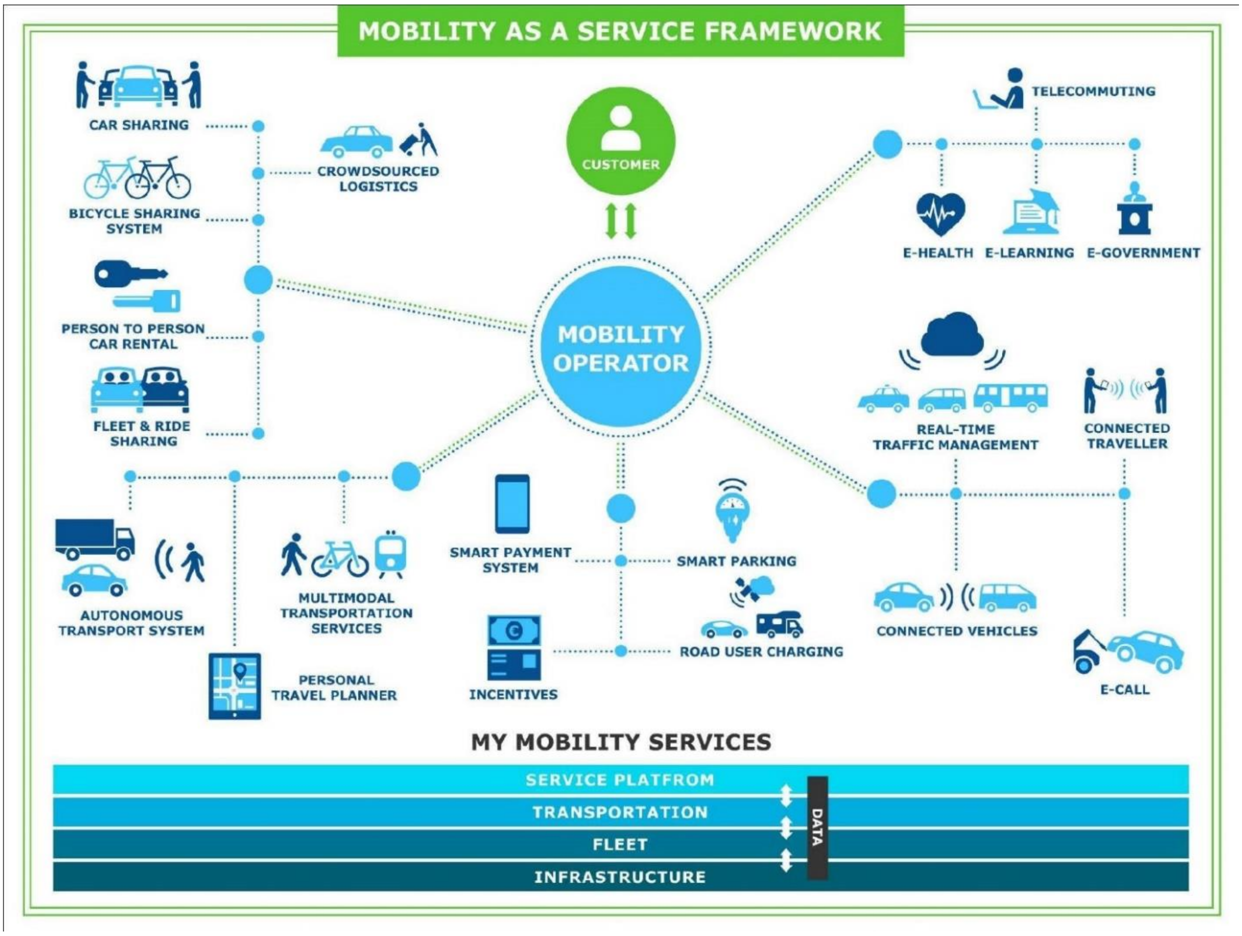



*Figure 11. The Mobility as a Service framework. Source: Reyes García, J. R., Lenz, G., Haveman, S. P., & Bonnema, G. M. (2020). State of the Art of Mobility as a Service (MaaS) Ecosystems and Architectures—An Overview of, and a Definition, Ecosystem and System Architecture for Electric Mobility as a Service (eMaaS). World Electric Vehicle Journal*

### *Data Sharing Model*

The MaaS ToMove project is structured as a decentralized system involving multiple levels of data exchange at local, regional, and national scales. Data flows between different actors, ensuring interoperability and enabling multimodal travel services. The core components of the data-sharing model include:

- **National Access Point** (NAP): a centralized repository of transport data used by MaaS operators to access real-time and scheduled transport information.
- **Regional Access Point** (RAP): managed by 5T, it aggregates transport data at the regional level and transmits it to the NAP.
- **MaaS Operators**: private entities that integrate transport services and provide a unified user experience. They are the only to have visibility over financial data (users' payment methods and transactions).
- **Municipality and Government Authorities**: entities which oversee data governance and leverage anonymised data for policymaking and analysis.
- **Incentives Platform**: technological tool collecting information concerning personal and service data in order to issue incentives for users based on different criteria (e.g. usage, different means of transportation used).

### Interoperability

MaaS ToMove operates as an interconnected ecosystem rather than a single platform, with multiple levels exhibiting a certain degree of Interoperability using API-based Data Exchange, which enables communication between MaaS Operators, transport providers, and access points.

There is **some extent of standardization**, not only at **data format** level i.e. public transport timetables data, but also in terms of **standardised contracts**: in fact, while private providers retain flexibility in service integration, **agreements with public transport operators follow a standard framework.**

### *Actors Involved*

Data Contributors:

- **Transport Service Providers (MSPs)**: supply real-time and scheduled transport data.
- **MaaS Operators**: serve as intermediaries, integrating transport services and sharing aggregated usage data with authorities.
- **Municipality** (5T as an Implementing Entity): manages governance and ensures data flow between stakeholders.

Users:

- **Citizens**: register for MaaS services via designated apps, receive incentives, and use multimodal travel planners.
- **Corporate Users**: employees of partner companies' access MaaS services with additional corporate incentives.
- **Tourists**: temporary users who receive travel benefits during big events (sport events, festivals and so on).

### *Examples of Shared Data*

The data shared within the MaaS ToMove ecosystem can be classified in the following main categories:

- **User Data**: demographic details required for registration and incentive allocation (handled securely by MaaS Operators).
- **Transport Data**: scheduled and real-time information from MSPs, including availability and pricing.
- **Usage Data**: travel history, mode of transport used, timestamps, and anonymized identifiers.
- **Payment Data**: managed exclusively by MaaS Operators, ensuring compliance with data protection regulations.
- **Environmental Impact Data** (currently work in progress): CO2 savings and other sustainability metrics, currently under development in collaboration with research institutions.

### GDPR Compliance and Data Processing

Given the handling of personal, transactional, and transport-related data, MaaS ToMove must adhere strictly to GDPR regulations. Key compliance measures include:

- Data anonymization: only anonymized data is shared with municipalities and governmental authorities (except those used to issue incentives, which are managed by the incentives platform and NOT by other public bodies).
- User consent mechanisms: users must explicitly consent to data processing when registering for services.
- Restricted access to sensitive data: payment information is exclusively managed by MaaS Operators, ensuring no third-party access.
- Privacy agreements: contracts between MaaS Operators, transport providers, and public authorities outline clear responsibilities for data handling and user privacy protection.

## Towards data spaces

Since Europe aims at creating a **single market for data**, a digital environment where data from various sources are integrated, analysed, and visualized to support informed decision-making processes, the tools and instruments previously described need to be oriented to this bigger scope. For this reason, **data spaces are considered core implementers of the European Data Strategy** and will thus be described in detail in the following chapter, from both a theoretical and a practical perspective.

To start approaching this subject *in an intuitive way*, **data spaces can be conceived as public squares where a market takes place**[9]. In a traditional market, the citizens establish an authority that regulates the market in their name, typically the municipality. Each farmer can participate and offer products in the market, if they follow the rules that the citizens have agreed upon and the municipality enforces in their name. The municipality will also enforce the basics of social interaction, such as preventing theft or antisocial behaviour. Similarly, in a data space, participants establish an authority that represents them (the data space governance authority) and participants can offer data and/or data related services (e.g. storage, processing, AI functionality) or consume the data and data services offered by others. Of course, a few participants will specialise in offering services, while others will use the data space just to purchase them. As in the case of a public square, it is rare that access is not open to everybody, as long as they follow the rules. Satisfying the participants' needs through data and services is not the responsibility of the governance authority, but rather of the other participants: the users ultimately decide if there is value for them participating in a data space, and the data space is expected to naturally evolve according to their needs, expectations and perspectives.

[9] European data spaces and the role of data.europa.eu - Publications Office of the EU

# Data Spaces

## A New Comprehensive Approach: Theory and Practice

Interreg Alpine Space
Co-funded by the European Union
Support EUSALP
EUSALP EU Strategy for the Alpine Region

# THE DATA SPACE APPROACH

In addition to the data architectures, tools, and methodologies that have been previously described and are in widespread use, a new solution is emerging that is rapidly sparking interest in the European context: data space. As data sharing requires collaboration and trust among a wide spectrum of actors, technology and service providers are developing instruments and tools to facilitate the exchange of information, while regulation on how to govern these instruments is rising. **Data spaces aim to reach the goal of pooling, sharing and reusing data – from both individuals and organizations – in a secure and privacy-preserving manner**. As part of the European strategy for data, this initiative supports Europe's global competitiveness and data sovereignty through the development of common European data spaces in various strategic sectors.

In the latest years, the construction and implementation of data spaces has already positively impacted private and public actors, by **approaching all data sets in a similar way**, **standardizing interfaces and connectors** – thus saving time and money, as well as increasing data quality – and **establishing conditions to access and share information**. In order to prevent the emergence of inequitable business models that would only benefit select few players, European regulations fostering data economy were defined, and both legal, business, and technical rules and technological standards (for interoperability issues and mechanisms to find data) have been developed to support these initiatives. A wide range of interpretations of this concept has spread. Indeed, the term "data space" is frequently employed to denote a more extensive category of simple data sharing systems. Nevertheless, this concept should be refined, as it pertains more precisely to **an approach to data sharing and value creation**.

According to the Data Space Support Centre (DSSC), a European Commission supporting initiative funded under the digital Europe programme, **a data space is an interoperable framework, based on common governance principles, standards, practices and enabling services, which enables trusted data transactions between participants**.

This definition highlights some crucial aspects of data spaces which will be further analysed in the following pages, while exploring this complex tool:

- **Interoperability**: shareable data need to be technically, semantically, organizationally and legally interoperable, thus data spaces establish principles and standards to follow in order to perceive both intra-data space and cross-data space interoperability.
- **Services**: the functionality of a data space depends on its services, which range from identifying data owners and consumers to making information findable and to allowing their re-use.
- **Trust**: participants need to trust the data space ecosystem to make informed decision about why, how and with whom share their data.
- **Sovereignty**: in a data space, individuals, organisations, and governments have control over their data and exercise their rights on the data, including its collection, storage, sharing, and use by others.

To approach the topic and orientate oneself in the structure of data spaces, a brief description of the fundamental concepts represented in the following diagram is offered.

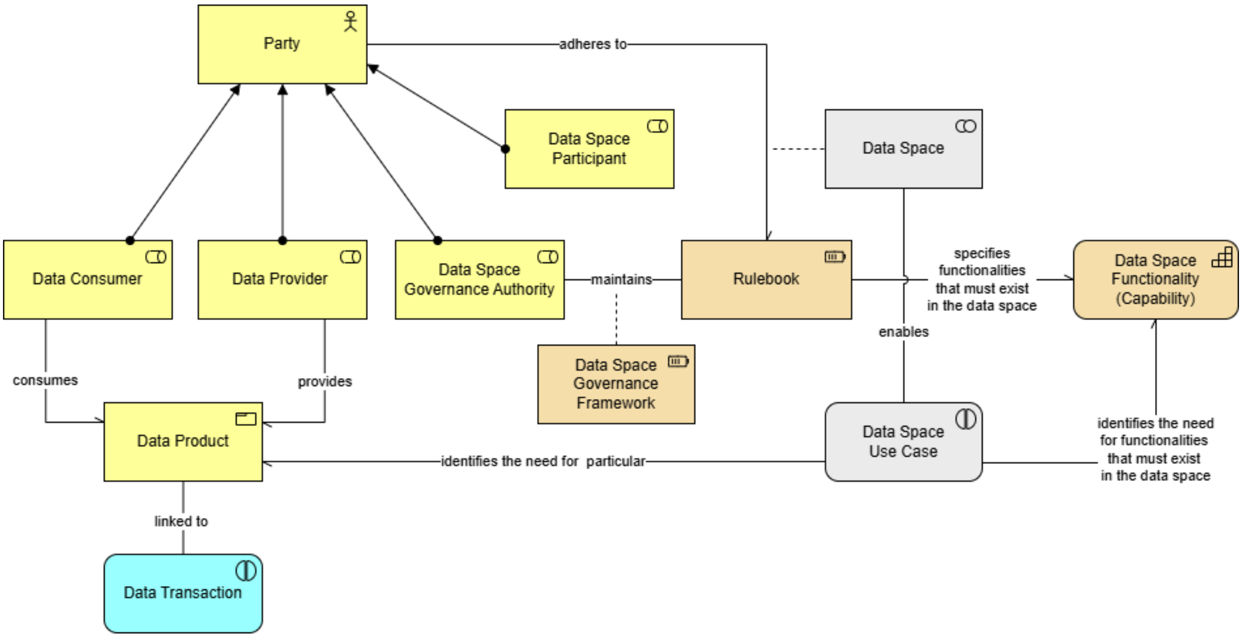



*Figure 12. Overview of key concepts. Source: DSCC Blueprint v2.0*

- **Participants**: entities participating in a data space may play different roles. They can thus be:
    - ***Data providers***, when supplying data to the ecosystem.
    - ***Data consumers***, when using data from the ecosystem.
    - ***Data space governance authorities***, when overseeing the data space.
    - ***Service providers***, when offering services to the data space.

- **Data space offerings**: data are the core of this ecosystem. As the objects to be shared, they are conceptualised as "**data products**" and may include the data itself, metadata, data usage policies, etc. A transaction of a data product primarily involves a data provider and a data consumer and may happen in various ways and across more and less complex scenarios, according to their sectors and to specific regulations governing those use cases.
- **Services**: necessary to store and process data, they are deployed by IT providers. Three classes of services may be identified:
    - ***Participant agent services***, required for an individual participant to join a data space (e.g. storing verifiable credentials, publishing data in a catalogue)
    - ***Federation services***, enabling the interplay of participants for all kinds of data sharing (e.g. providing a shared list of available participants and data products, sharing policy information and regulations eventually synergic to specific EU legislation)
    - ***Value-creation services***, to support value creation from data available within the governance framework (e.g. tools for analysis, tracking and tracing functionalities)
- **Data governance frameworks and rulebooks**: since participants can make their own decisions on exchanges in the ecosystem, "data sovereignty" is one of the key features of a data space. Nevertheless, a governance framework is needed, supporting autonomy of providers and consumers, as well as defining common rules and regulations on services which apply to the whole spectrum of actors involved. A specific rulebook collects each data space regulations and is managed by the data space governance authority.

The main features of data spaces make them particularly interesting from an economic point of view, as they may foster a comprehensive approach to the growth of the Alpine regions in EUSALP as a whole. To this end, according to European standards, taxonomies and vocabularies, this ecosystem will be thus described following the fundamental pillars identified in the latest version of the Data Space Support Centre Blueprint, which aims to support data space initiatives by deconstructing them into interrelated building blocks, facilitating interoperability and making it easier for providers and consumers to participate in more data spaces. The architecture which will thus be described aligns to the features of Gaia-X and IDSA's ones, as the main initiatives focused on conceptualisation and standardisation of data space experiences.

The development of the infrastructure of a data space is achieved through the implementation of 17 building blocks, grouped in 6 pillars and divided into 2 main classes, as shown in the following schema. Business and organizational category includes business, governance and legal building blocks, while technical category includes data interoperability, data sovereignty & trust and data value creation enablers.

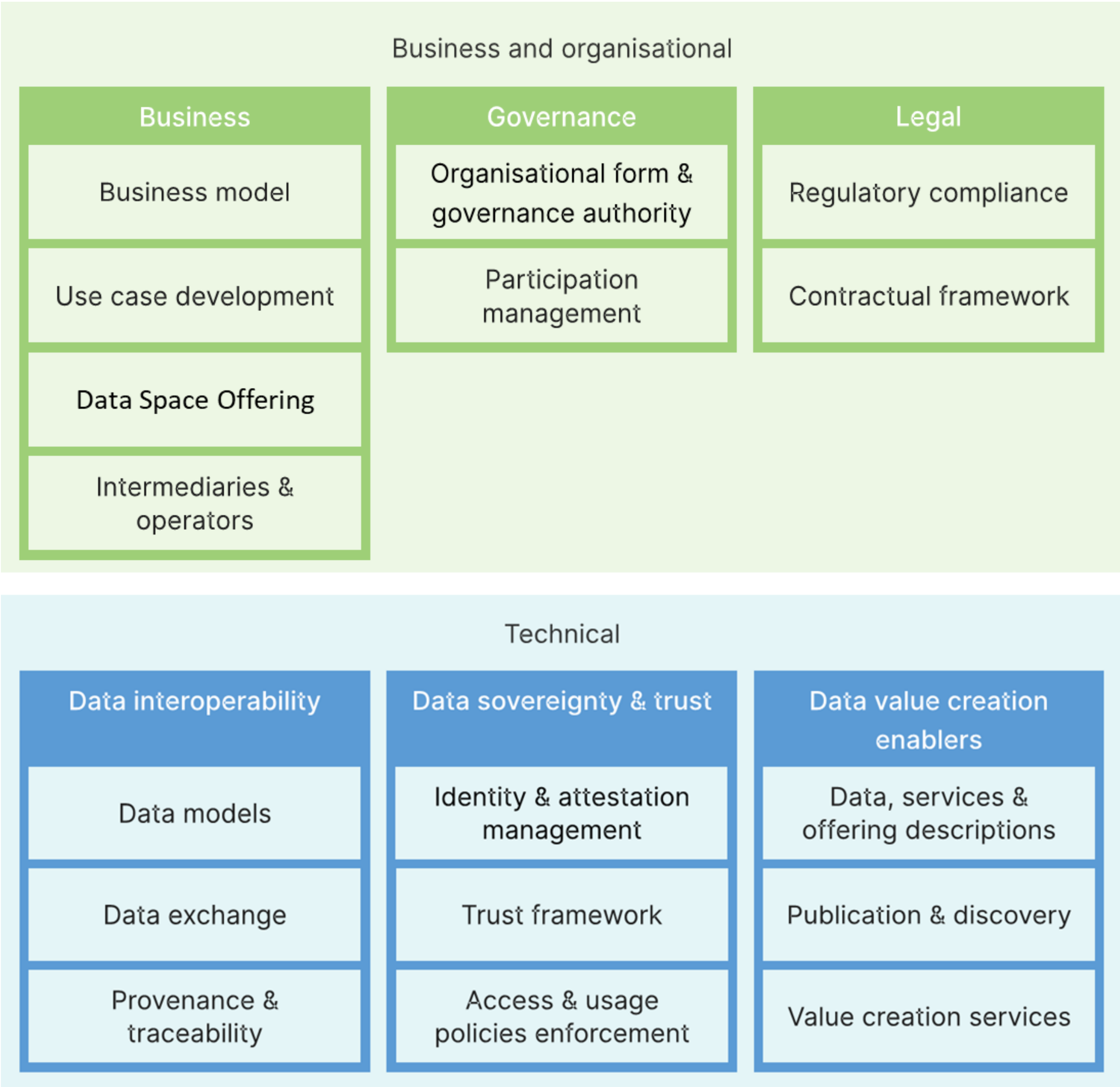
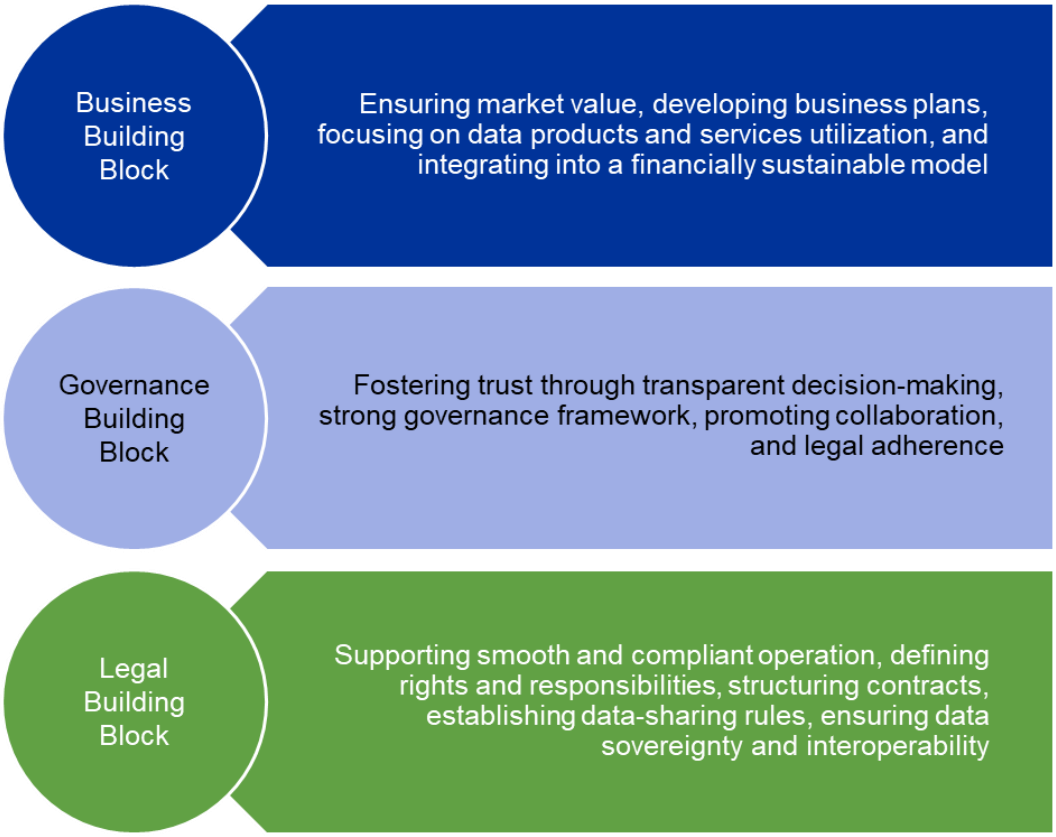



*Figure 13. Building Block overview. Source: DSCC Blueprint v2.0*

## Business and Organisational Building Blocks

As mentioned above, a data space is made up of technology as well as of business, governance and legal aspects, which are addressed in this first class of building blocks. This section therefore has a threefold objective:




Business and Organisational Building Blocks show the **multi-layered structure of a data space**: the **participant level** concerns incentives and behaviours of individual data space participants and is the basis for the **use case level**, in which at least two participants create value from data sharing. The proper **data space level focuses** on scopes and boundaries established by the governance authority and interacting with the **broader** economic, legislative and societal **ecosystem**.

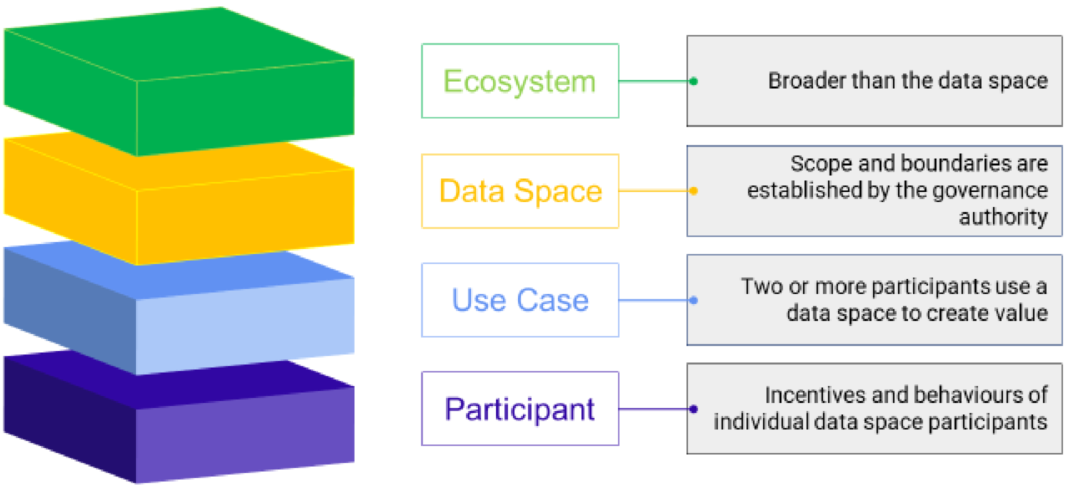



*Figure 14. Levels of data space business and governance. Source: DSSC Blueprint v2.0*

## Business Building Blocks

The following four building blocks aim to **provide guidelines for strategic decision**. In fact, this pillar focuses on the relationship between data products and services in order to design and enforce an efficient and sustainable business model and find winning strategies to interact with stakeholders and make the data space valuable.

### *Business model*

A data space requires collaboration between multiple actors: data providers, data consumers and service providers should be able to generate value from this solution, agree on how actors share data as well as on how to balance costs and revenues. Thus, a data space needs to **combine concepts from different business models to successfully develop and maintain this ecosystem**. The value of a data space is to **enable sharing and reusing of data in context of use cases**, by establishing efficiency because of standardised interfaces, and sovereignty, through control over data sharing.

When applied to a data space, key concepts of a business model are as follows:

- **Value proposition**: in a data space there must be multiple value propositions for the different actors involved. Each participant may always benefit from the total offering of other participants (in terms of data, metadata and related services) and the way in which the data space is governed.
- **Multi-sidedness**: as a data space serves interaction between participants of various types, its business model should outline the benefits, conditions and costs for each category. The term "multi-sided" also means that the value for data providers (supply side) increases as more consumers (demand side) participate in the data space, and vice versa, thereby enforcing the appeal and attractivity of the whole ecosystem – the so-called "network effect".
- **Collaboration**: the business model of a data space is collaborative since it applies to a set of public and private entities, and value can be generated only when models of each participant are coherent with those of others.
- **Governance authorities**: these bodies are responsible for monitoring access conditions and sharing operations within a data space, and ensuring it remains profitable for all the actors involved. These regulations need to be aligned with the entrepreneurial objectives of the ecosystem.
- **Objectives and principles**: strategic decisions on the organisational form of the data space, the targeted operational scale and the alignment to principles of fairness, sovereignty, transparency and human-centricity influence the value propositions and thus the business model of data spaces.

To outline the capabilities and benefits of the data space, value propositions are established for all stakeholders. A crucial role is played by objectives and principles in attracting all parties. A collaborative business model approach is essential to develop the data space in harmony with the various actors, and the governance authority is tasked with continuously ensuring that the needs of all these stakeholders are met.

## *Use case development*

Data space use cases are **settings where two or more participants create economic, social or environmental value from data sharing**. For this reason, they are extremely important in the early phases for **attracting users and providers** and **supporting the growth of the data space**, and they need to be updated and improved as their business environment evolves.

Use cases can range from being very simple to very complex situations, with multiple parties involved in the exchange of data products and related services. However, at the heart of a use case there is a **transaction from a provider to a recipient**. This is true even when dealing with cross data space use cases, where participants of two or more data spaces jointly create value from data sharing. The need for orchestration is obviously increased in situations with a high number of parties. For this reason, in the context of the governance framework, each use case is subject to the oversight of an **orchestrator**, which is entrusted with the responsibility of regulating the use case and aligning with the data space governance authority.

The development process is iterative, aiming at shaping a scenario, refining it and realising the use case. Four steps may thus be identified:

1. **Identifying and monitoring use case scenarios:** in this phase potentially valuable use cases that require data sharing are designed, starting from the needs of current and potential data space participants or from already existing sharing contexts, data products and value creation services which may benefit from data space features. The list of identified scenarios will thus be prioritised and updated, mapping their development stage, reasons for their potential success, or for their abandonment, and discussing potential collaboration with other data spaces. Before refinement, orchestrators need to have been identified, for the governance authority to have an owner of the following development work.
2. **Refining use case scenarios:** in this step, the orchestrator refines the use case scenario, managing all the challenges leading to its implementation. In case of complex scenarios, refinement is an iterative co-creation and design process. Purposes, value propositions, data products and value creation services needed should be discussed and agreed by all the parties involved in the use case. Moreover, business model, regulation, contractual issues, security and interoperability are addressed, as well as a plan of implementation activities, which establishes the sequence in which different parts of the use case will be executed.
3. **Implementing use cases:** having all the preconditions met (including an operational data space infrastructure, the use case participants joining, data products and value creation services available), the use case scenario is implemented and starts creating value as previously planned. The orchestrator is responsible for organisational and technical features to be developed.
4. **Continuous improvement process:** due to evolutions in technologies, data products, participants and regulations – as well as to improve the performance, use cases are constantly developed, carefully managing the challenges met. A use case life cycle continues until it is removed from the market. It may grow in complexity, wideness of products and services available, attractivity towards new entities, but it may also decline whereas it does not reach sufficient adoption, or it reveals not to be financially viable. Changes will then occur, and the orchestrator will prioritise those which are expected to lead better results.

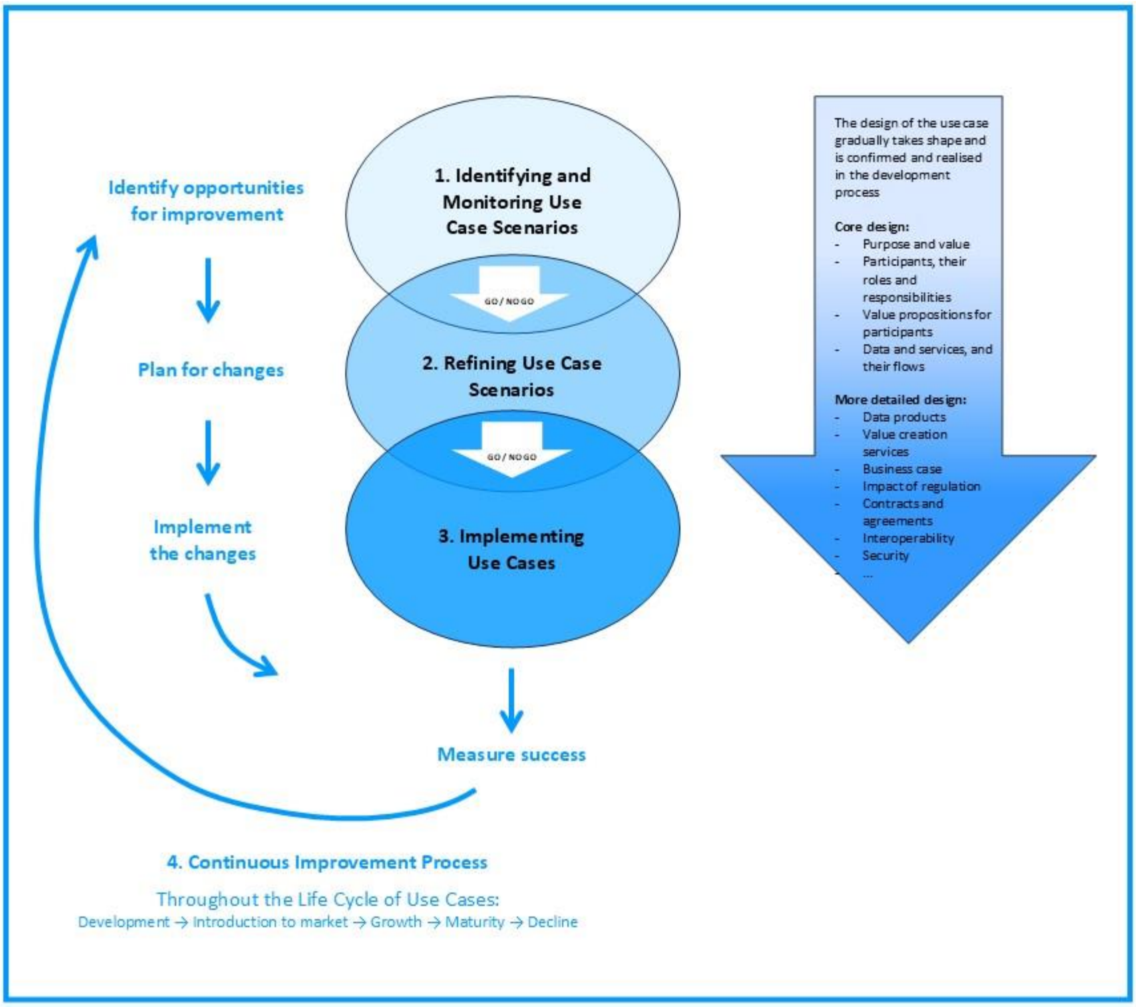



*Figure 15. Elements of Use Case Development. Source: DSCC Blueprint v2.0*

## Data space offering

The data space offering groups all the **offerings available to data space participants**, in terms of **data products and services** available to participants, accompanied by their **descriptions** to facilitate informed decision-making on potential exchanges. This building block thus proposes a strategy for offerings identification, management and governance, and it is extremely important for the data space governance authorities, which are responsible for the data product and services offered, their regulation, the design of attractive offerings fostering network effects among participants through the use cases implemented.

**The offerings of data products and value creation services are collected in a catalogue to make them discoverable by participants**. **Data products are assets that provide monetary and/or non-monetary value form data, which have been previously transformed – together with metadata and associated licence terms – into consumable and marketable objects**. Whatever is the thematic content of the product, it typically includes information about quality, format, frequency, duration, access and control rights, delivery options, provenance, and pricing. Moreover, data products are recommended to adhere to the FAIR principles (findability, accessibility, interoperability and reusability). Providers hold the responsibility for data products over their lifecycle, including maintenance and the respect of terms and condition of both the data space ecosystem and the data rights holders. On the other hand, the offering may also include value creation services, e.g. data visualization, anonymization, quality assessment, connection to external infrastructures.

From a strategic perspective, a data space offering develops around these three elements:

- **Identification and onboarding of priority data products and services that serve existing and future use cases:** data products and services aim at creating value within use cases. Identifying the assets that can provide value across multiple use cases is essential to support the growth of the data space.
- **Development, maintenance and enforcement of the governance rules of the data space offering:** the governance authority sets the rules for the data products and services and thus enforces trust towards the data space, by ensuring sustainability and coherence to principles of quality trustworthiness, security, privacy, interoperability and ethical considerations.
- **Supporting the participants to develop and offer high-quality data products:** while creating measurable value, data products improve the utility of the data by bundling policies, business and contractual information with data, foster synergies both intra and cross-data spaces, attracting new participants, and define common rules which ensure stronger compliance and greater transparency. For these reasons, the governance authority should support participants in creating data products, also by providing tools and processes and promoting data owners to develop multiple products with various options of the associated information, to reuse them in a potential wider range of use cases.

## *Intermediaries and operators*

This building block focuses on **what kind of business, governance, legal, and contractual topics the data space governance authority should focus on while procuring enable services from one or more operators** (usually referred to as "**intermediaries**"). In the data space ecosystem, those actors are responsible for providing enabling services, thus their effectiveness can be assessed by their ability to streamline data sharing, improve data space accessibility and usability, support scalability and interoperability intra and cross-data space. When facing questions related to intermediaries, the governance authority should also consider risks associated with procurement, e.g. vendor lock-in, additional costs, and potential loss of sovereignty on the data space.

To better understand the strategic relevance of intermediaries, it may be useful notice that services are – together with data – the core of a data space, and whether they are technical, business or organisational, they enable the trusted data sharing which characterises this ecosystem. **A service provider** is a business entity that provides this kind of services: **depending on the governance authority, it may be considered as a participant of the data space** (e.g. when providing catalogues for the offerings) **or not, and will thus be subject to various requirements, limitations and responsibilities, basing on different levels of engagement in the data space.** Within this broader category, **an operator or more intermediaries are specifically responsible for enabling data sharing and trusted transactions**: those entities typically provide neutral and trusted federation and participant agent services but also try to increase accessibility to the data space, enable interoperability and offer business and organisational services which may be synergic to the value creation ones.

The governance framework should include information regarding all service providers in the data space, describing which kind of service is provided, whether the governance authority or individual participants pay for the service, and whether the contractor is also the final user or not. Allowing independent service providers to operate in the data space may be more attractive; on the other hand, by enforcing those entities to commit to the data space rulebook the governance authority regulates service provision in a stronger way. The governance framework designs how intermediaries provide value and how risks are managed, thus defining different kinds of rights and responsibilities and ensuring service compliance with principles of neutrality and scalability. Moreover, service providers are also subject to broader legal frameworks, e.g. competition law, intellectual property law, data protection, digital services regulations and compliance, sector-specific, regional and national regulations. To reduce risk of losing sovereignty when dealing with a single operator, at least standards and adoption of open-source technologies may be required.

Since service providers contribute to the overall economics of the data space, interoperability and collaboration are to be ensured and enhanced. Intermediaries can cooperate with each other both when providing different and complementary enabling services, and when furnishing the same service, by adopting the governance authority’s standards and frameworks, thus preserving compatibility. Even when competing to provide the same services to different ecosystems, intermediaries and operators try to setup a fungible and interchangeable service offering, in order to facilitate cross-data space interoperability.

## Governance Building Blocks

The following two building blocks aim to **help in shaping the structure, decision-making processes and participant engagement within data spaces**, by providing a holistic framework for secure and efficient operations.

### *Organisational Form and Governance Authority*

This building block explores **different legal structures for a data space** – which can be both an unincorporated or an incorporated entity – and focuses on how to **define the governance authority and its related framework.** The **governance framework** is a **set of internal rules and policies applicable to all data space participants**. In particular, the organisational form refers to the type of legal entity the data space can assume, while the governance authority is responsible for developing, applying and maintaining the internal rules.

It is crucial to define the organisational form and governance authority early in the development process for ensuring the effective operation of the data space, fostering trust and ensuring data quality. They should be preferably established before entering the operational phase and be flexible enough to both follow and guide the data space evolution over time, as it grows and includes new functions. Governance framework includes functions such as conflict resolution, access control, and risk management. The choice of legal form for a data space impacts the type of governance and the ability to enforce internal rules.

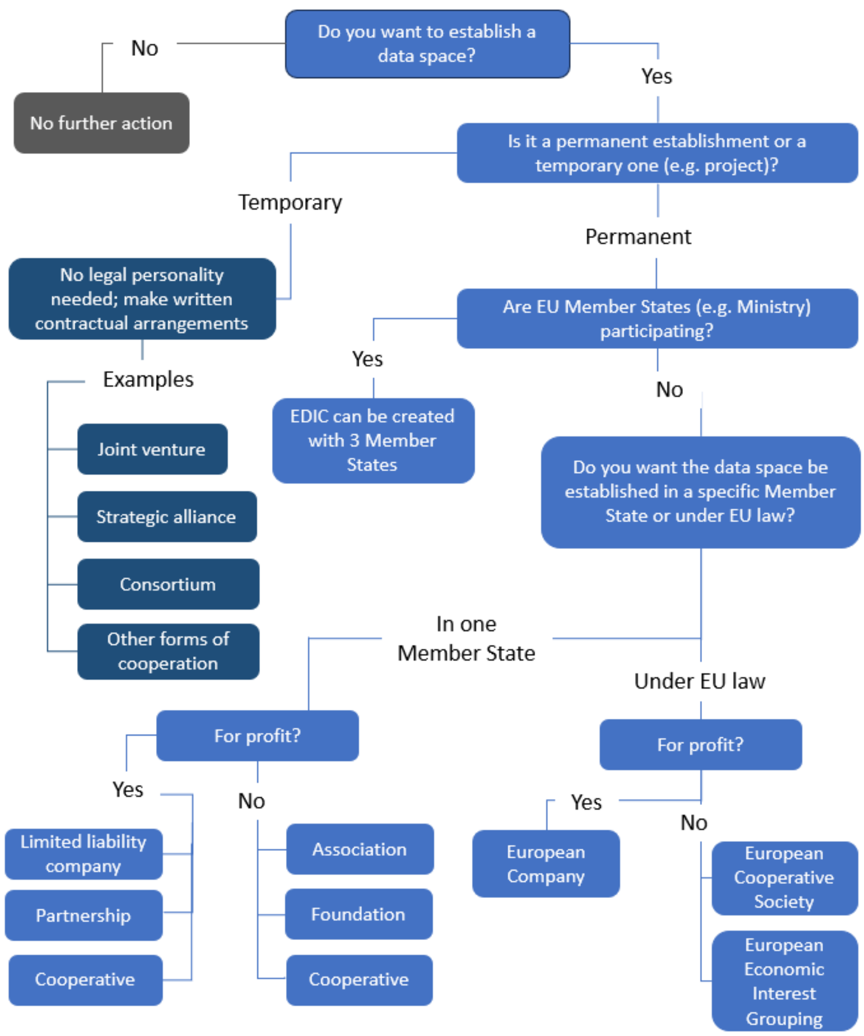



*Figure 16. Decision tree including some of the core design choices for organisational form. Source: DSCC Blueprint v2.0*

A data space usually originates from **data sharing needs of several organisations**. The members must reach several basic agreements (e.g. on the desire to establish a data space, targeted economic sectors, collaborative approach to achieve this objective). Key aspects to be discusses may thus concern the type of the data space establishment (permanent or temporary), the legal personality, whether the data space aims to generate profits, the level of involvement of the members, and the headquarters' country.

If there is uncertainty about the permanent establishment or disagreement about the legal form at the beginning of a data space initiative, the committed members can use various contractual arrangements to give structure to the data space as a multi-year project. The result of this cooperation is an **unincorporated data space**, not having its own resources but **relying only on the resources, capacities and capabilities of its members**. In these cases, it is advisable to expressly allocate distinct roles and tasks in the contract, in order to minimise risks of scope gaps and potential internal disagreements. An unincorporated entity must determine which member will be responsible for data security according to GDPR (even when the data space is not focused on handling personal data) and must also designate a representative for interactions with other data space participants, including transaction participants, service providers, and third parties. Being managed by the members, unincorporated data spaces must establish strong governance mechanisms to coordinate activities and achieve common goals. Although governance does not extend to managing resources or outcomes, members can set up a governance authority, called "general assembly", to make decisions and reach cooperation's goals. However, an unincorporated data space is unlikely to be sustainable in the long term and will eventually need to become incorporated to develop independently from its members resources.

When a data space is established as a permanent structure, it is necessary to create a new legal entity or base on an existing one. Not-for-profit data spaces are often set up as associations or foundations, while limited liability companies are the most common structure for profit-driven experiences. If members of a data space initiative decide to establish it under national law, they can choose from various legal forms across the 27 EU Member States, each with its advantages and disadvantages. The following European legal forms are especially attractive as they allow companies to operate as a single entity throughout Europe:

- **European Digital Infrastructure Consortium**, a special legal form created for Common European Data Spaces and special multi-country projects.
- **European Company**, a for-profit legal entity, like the limited liability company.
- **European Cooperative Society**, a non-profit with characteristics of a cooperative and public limited company.
- **European Economic Interest Grouping**, a non-profit legal entity like a partnership.

Once the legal form has been chosen, members sign a founding agreement and undertake to specific rights and obligations. An **incorporated data space can enter contractual relationships with other legal entities, be beneficiary of legal rights, be responsible and liable for its obligations and employ own staff**. It requires bodies to manage operations and enforce internal rules. The composition and competencies of these bodies are determined by the founding members and depend on the legal form and size of the data space. Moreover, the members decide on the competencies of the governance authority, which can range from small structures to complex entities with specialized bodies, as seen in large common European data spaces. If the data space is based on an existing legal entity, it may be useful to create a separate body for management.

According to the governance framework, each data space is governed by external (legal) and internal (decided by members) rules, both collected in a rulebook tailored to the data space specific needs. The founding agreement defines the internal organization and decisions, while internal documents, including regulations and technical specifications, must be approved by the decision-making body.

The rulebook ensures flexibility and compliance and should be available in both human-readable and machine-readable formats.

## Participation management

This building block addresses **governance processes for managing participant engagement in data spaces**, thus including participants' identification, onboarding, offboarding, and the establishment of rules for data transactions and service provision. It addresses risks like data governance challenges and reduced collaboration. Moreover, this building block provides guidelines for efficient and secure participation by integrating relationships with other governance aspects like regulatory compliance and identity management.

As previously stated, participants include organizations and individuals who, according to the different roles allowed by each data space governance rulebook, engage in data transactions or act as data rights holders. This building block emphasizes that **successful participation depends on a shared mission**, and its management is essential for the proper functioning and governance of data spaces. A participant lifecycle may be described as follows:

- **Onboarding:** efficient onboarding of participants should ensure quick integration and compliance with regulatory and technical standards. The governance authority sets the minimum requirements for participation, including the general terms and conditions that define admission policies and compliance standards (e.g. identity verification, technical and security requirements, data protection regulations, such as the GDPR). The onboarding process must thus verify entry conditions and ensure technical interoperability after integration. The governance authority manages participant admission and continuously checks compliance with the data space's rules and obligations. Participant feedback is crucial for improving the process and ensuring long-term security and satisfaction.
- **Data Transaction:** in the operational and scaling phases of a data space, the number of participants and use cases grows. The governance authority facilitates interaction among participants and must address imbalances between supply and demand, while attracting new participants. As the data space expands, the authority must also adapt the governance framework to meet new needs, such as regulatory changes or the inclusion of new sectors or countries. Therefore, the governance framework must remain flexible and inclusive, ensuring compliance, security, and interoperability, while staying aligned with the central mission of the data space.
- **Offboarding:** it typically consists of a participant wishing to exit from the data space and following a structured offboarding process. In other cases, the exit may be forced, for example, due to non-compliance with the data space rules or to exceptional situations, such as the bankruptcy of a participant. The governance authority must inform all affected parties and enforce the exit, with the governance framework ensuring a smooth transition and protecting the interests of all involved parties. When leaving a data space, all participant's contracts must have been fulfilled and no open obligations should remain. Key points that must be followed to make the offboarding process efficient concern:
    - Documentation of the exit procedure that ensures clear, consistent steps for data transfer, access termination and contract closure.
    - Protocols for safe data transfer and/or deletion.
    - Notification to inform all parties promptly, avoiding misunderstandings.
    - Verification of compliance to confirm all obligations are met before completing the exit.
    - Offboarding support to provide assistance during the transition period.
    - Periodic governance framework reviews for its updating to address regulatory changes.

## Legal Building Blocks

The following two building blocks focus on regulatory compliance and contractual framework, aiming to **enhance understanding of compliance and contract issues within the data space**. Since this kind of requirements varies across sectors and depends on business, governance and technical infrastructure decisions, basic and common features will be outlined.

### *Regulatory Compliance*

The data space governance authority is responsible for applying legal rules to the ecosystem design and operation. This building block supports the **establishment of internal policies and responsibilities** – according to the role of a data space participant in the ecosystem as a whole and in specific data transactions – and the **monitoring of regulatory compliance**, which is an ongoing practise throughout the data space lifecycle.

The following sector-agnostic mapping of legislation that could apply to data spaces, their participants, the data and services offered has recently been deployed by the DSSC. However, each sector and each use case may be subjected to different legal frameworks, and regional and national legislations should be further analysed.

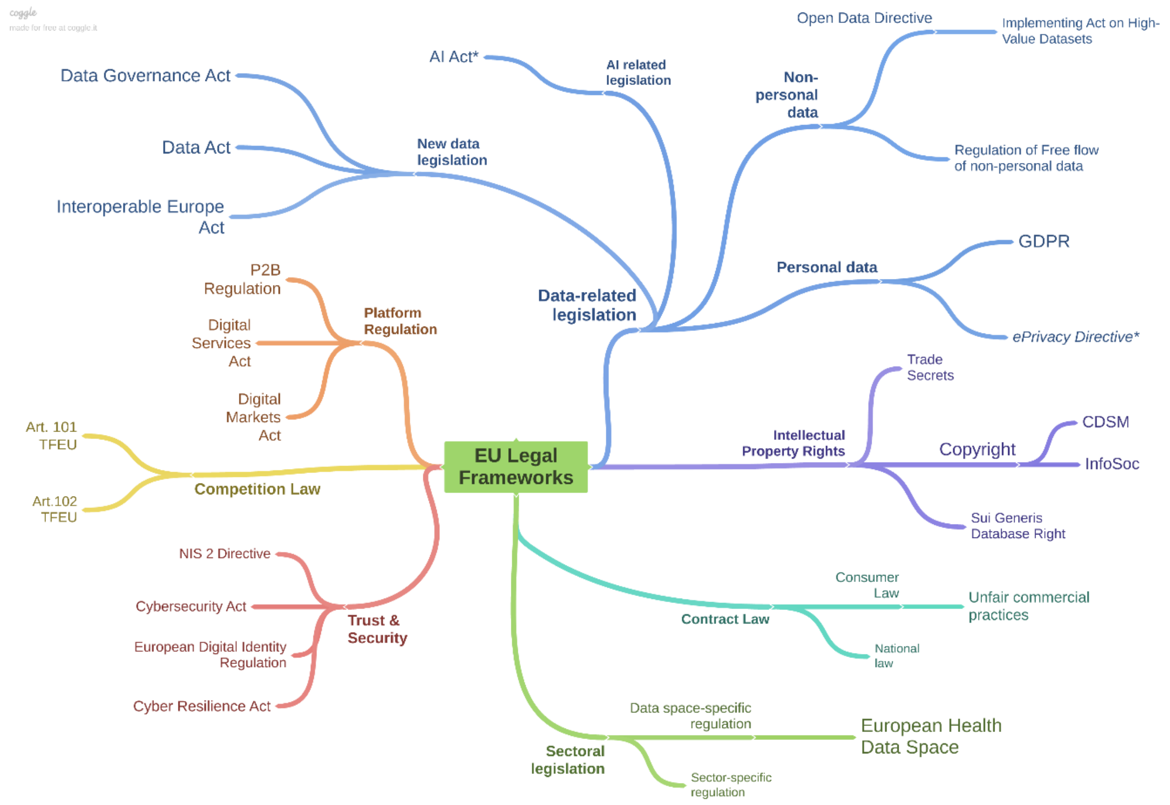



*Figure 17. EU Legal Frameworks applicable to data spaces development and operation. Source: DSCC Blueprint v2.0*

The following five interrelated elements are addressed by this building block:

- **Triggers**: a trigger is an element, criterion or event occurred in a particular context of a data space signalling that a specific legal framework must or should be applied. Triggers may pertain to types of data (kinds of data available in a specific context), data space participants (legal status of participants), or use cases (domains and sectors):

    - When dealing with certain categories of data, particular protection regulations may impose restrictions on access and processing (e.g. GDPR protects natural persons on the processing of personal data, as intellectual property and trade secrets do for other kinds of data).
    - Data space participants can be conventionally divided into private and public entities, but it is necessary to remember that specific rights and obligations may always be imposed, depending on whether an entity operates as a public sector organisation, a private company, or a consumer.
    - Specific sectoral regulations may impact the development of a particular use case and should thus be monitored together with the cross-sectoral legal frameworks.

- **Data space requirement:** this category groups legislations that directly regulate data space. Among these requirements, the Data Act defines essential criteria concerning interoperability and data-sharing mechanisms.
- **Additional legal considerations:** these are complementary requirements that a data space should fulfil, e.g. cybersecurity legislative frameworks to protect data integrity and privacy.
- **Tools enabling regulatory compliance in the data space:** together with organisational and contractual measures, technical tools may help addressing the complexity of a data space ecosystem and supporting compliance within it.
- **Regulatory compliance flowcharts:** decision trees[10] can facilitate the identification of relevant triggers and the setup of a checklist of questions by both the governance authority and the participants. This process enables effective management of potential issues detected.

## *Contractual Framework*

This building block describes the **legally enforceable agreements that regulate operations and relationships among parties within a data space**. It thus supports the governance authority in translating legally relevant elements (e.g. responsibilities, liability, rules on onboarding and offboarding) into **interoperable, automated and scalable agreements**. In coordination with the Regulatory Compliance building block, this framework aims to distinguish mandatory elements from the negotiable ones.

Agreements are physical or digital documents that serve as evidence of a contract, a binding among parties with a common interest in mind that creates mutual obligations enforceable by law. Although they can vary according to various elements affecting their structure and / or their content, they always regulate relationships and contractual rights and duties among data space participants, members and enabling service providers. **This framework guarantees to data providers and data rights holders both data sovereignty and the possibility to add specific conditions to the baseline governance established by the authorities.**

Three main categories of agreements can be distinguished:

- **Institutional agreements** underlie the creation of data spaces and provide a minimum mandatory governance framework which applies to all participants. These agreements can be, at a high level, a common set of rules that regulate the relationship among data space participants or, at a more granular level, a set of specific terms and conditions for data product

[10] Examples of decision trees are illustrated in Regulatory Compliance section of the Data Space Support Centre Blueprint

contracts. They thus allow the introduction of common elements across the data space, giving legal effect to the organisational and business decisions applicable to the data space.

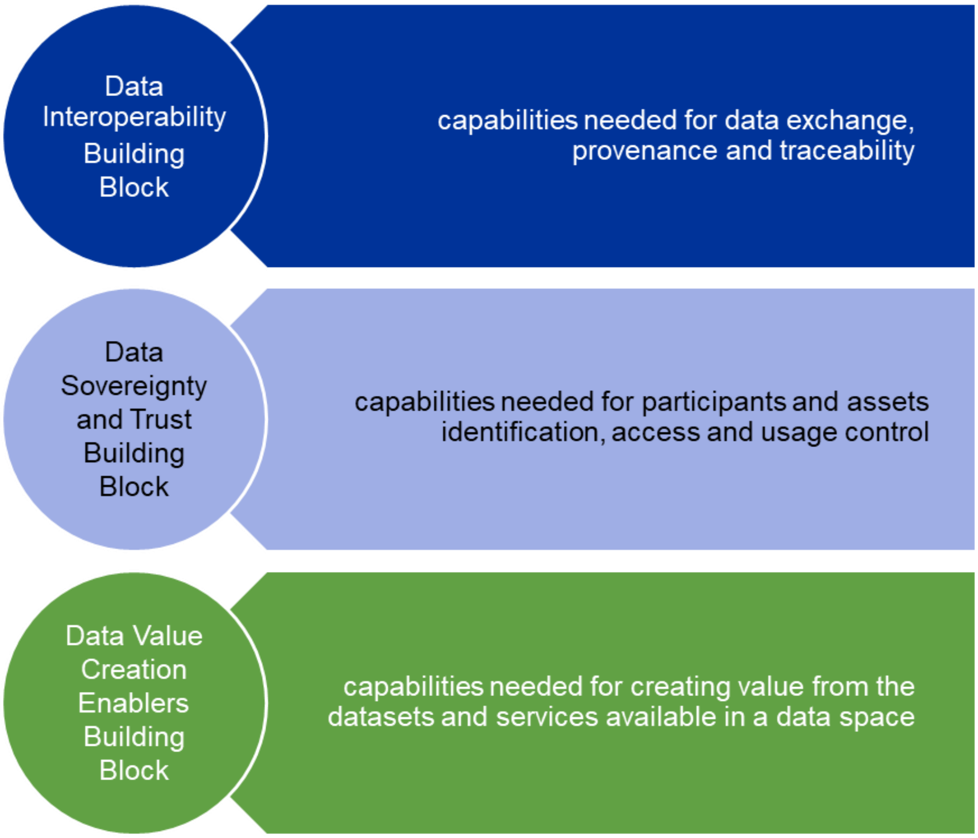
Among institutional agreements:

- ***Founding agreement*** establishes the data space and it sets out the purpose of the initiative as defined by the founding members, the general responsibilities and liability, and the reference to the chosen governance models. When the aim is to set up a data space as a separate legal entity, articles of association or statutes should be preferred; on the other hand, cooperation agreements (including consortium agreements and joint ventures) establish an incorporated data space. Moreover, these different organisational forms define different ways for a new entity to become member of the data space.
- ***General terms & conditions*** contribute to providing legal enforceability to some elements of the governance framework, by defining roles and responsibilities for the participants. Concerning admission policy, intellectual property policy, data protection policy, technical standards, cybersecurity and risk management policies, complaints policy and dispute resolution rules, these agreements mainly affect data transactions and interactions between providers and consumers.
- ***Agreements related to enabling services*** concern the provision of the services necessary for the data space to work (e.g. data-related services, agreements for the provision of trust framework services, agreements for the management of identities).

- **Data-sharing agreements**, which only involve data transaction participants, as concerning supply and transfer of data. Terms and conditions that govern the provision of a data product are core in such agreements, which are defined by data providers and thus reflect the principle of data sovereignty. A specific type of data-sharing agreement is the data product contract. When designing data products, some legal aspects should be considered: Lawful source of data: when data content is copyright-protected, the data space may be liable for copyright infringement. Contractual clauses (e.g. warranties and indemnities) address the risks associated with this liability.
  - **Applicability of sectoral regulation and/or horizontal regulations**.
  - **Jurisdiction and applicable law**: alignment between data product contracts and the general terms and conditions of the data space is mandatory, but each contract must include specific clauses that regulate which law applies to it and which courts have jurisdiction over potential disputes.
  - **Smart contracts**: they enforce legal obligations through technical means, thus enhancing trust in a data transaction and reducing risks of contractual breaches. They can be used to formalise legal agreements, negotiations and execution, and serve to prevent unauthorized access to data. In a data space, these contracts must adhere to Art 33 and Art 36 of the Data Act, facilitating the interoperability of tools that automate the execution of data-sharing agreements and defining essential requirements of access control, safe termination and robustness for smart contracts.
- **Services agreements** relate to all services that have data sharing as the subject matter of the agreement (e.g. data processing, data marketplace contracts, personal information management systems) or to services for either individual data space participants or the data space as a whole, aimed at enabling functionalities in the data space (e.g. federation services, participant agent services, value creation services).

## Technical Building Blocks

In addition to business, governance and legal aspects, technical capabilities are also needed in a data space. Providing standard specifications for these enables governance authorities and participants to make more informed technical decisions and supports the reuse of technical solutions and interoperability between different ecosystems. This second class of building blocks thus addresses:




The technical building blocks provide scope, specifications and standards for these capabilities, which can be implemented though services by specific providers with the necessary software components.

## Data Interoperability

The following three building blocks aim to provide a **clear understanding of data**, which is necessary for the proper operation of a data space as an ecosystem of participants, products and services. This should be achieved both on a semantic and a technical level. While semantic interoperability focuses on the meaning of concepts, technical interoperability is concerned with a specific syntax. Data spaces must define data models and standardise the technical interfaces (APIs) for data exchange. Furthermore, the execution of the data-sharing process might require tracking for the purposes of provenance or traceability to make the process auditable.

### *Data Models*

A data model is a **structured representation of data elements and relationships used to facilitate semantic interoperability within and across domains**, as they ensure accurate and consistent interpretation of data exchanged in a data space. Composed of metadata describing semantics, **these models allow data providers to offer data products and data consumers to use them**. The adoption of the same data model enables semantic interoperability, facilitating data exchange among the ecosystem participants. In fact, they set up "semantic standard" and serve as dictionaries allowing providers and consumers to "speak the same language" when exchanging data. Nevertheless, it is important to note that the pursuit of a balance between the need for strict uniformity to keep data consistent and easy to understand, and the need to accommodate the divergent data requirements of different organizations, is an ongoing process.

To properly deal with data models, a data space should include the following capabilities:

- **Data model development**, concerning reuse or creation of models to ensure uniformity and interoperability.
- **Data model governance**, related to management and maintenance of models, ensuring them wide consensus among the data space participants.
- **Data model integration**, by linking each offering to a model describing its structure and semantics.
- **Data model abstraction**, to manage the transition from semantic to technical interoperability.
- **Data models across data spaces** to promote semantic interoperability among different ecosystems through standardised discovery of models.

The models for each data space are stored in a Vocabulary Service, a common repository to which both providers and consumers can refer during an exchange, since each data product relies on a data model in this vocabulary. However, this is a challenge for federated data space as they should access other data spaces to reuse their models. For this reason, data spaces must ensure that data offerings are linked to models that describe their structure and semantics, using existing standards or extending them if necessary (thanks to data space intermediaries).

Different abstraction layers in data models can be distinguished, with each layer consisting of metadata about the shared data:

- **Vocabulary**: basic concepts and relationships expressed as terms and definitions.
- **Ontology**: knowledge by modelling information objects and their relationships.
- **Application Profile**: data model for applications that fulfil a particular use case.
- **Data Schema**: data exchange technology specific representation of application profile, including the syntax, structure, data types, and constraints for a data exchange.

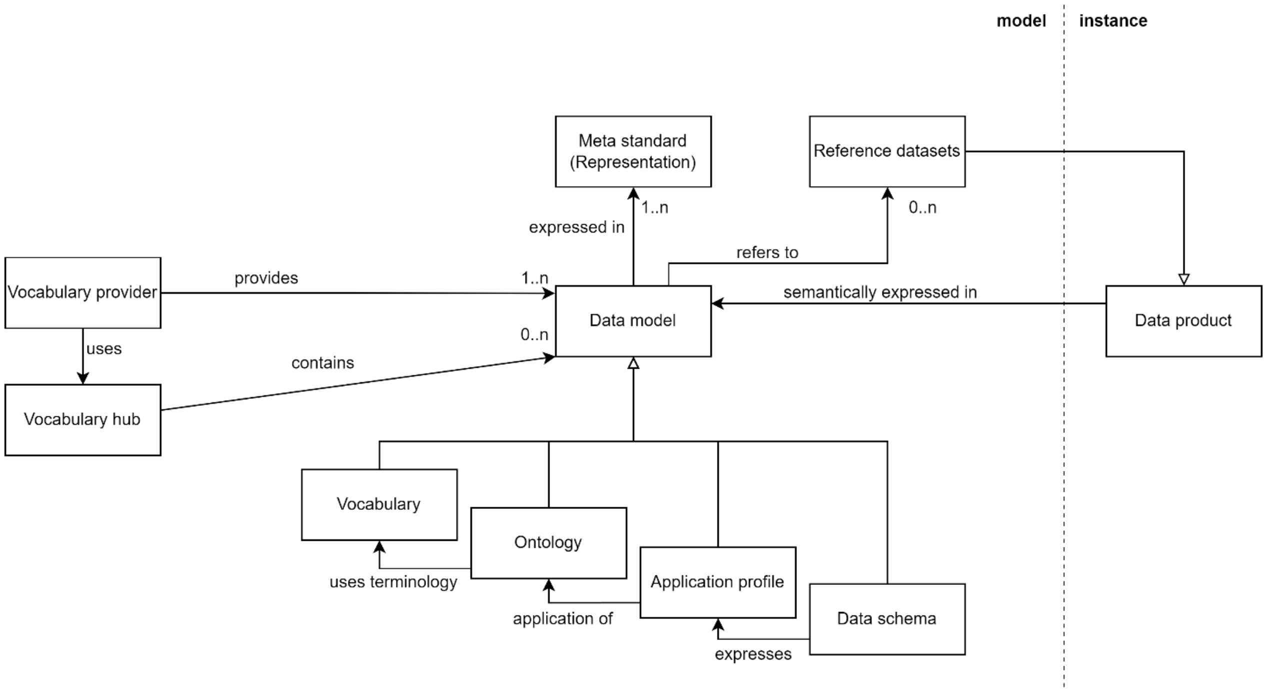



*Figure 18. Conceptual model of the Data Models Building Block. Source: [DSCC Blueprint v2.0]*

In a data exchange protocol, a **data schema** defines the **structure and format of data exchanged between clients and servers**. This schema outlines the properties, types, and constraints of data fields, thus ensuring consistency and effectiveness in the exchange. To support interoperability, data spaces should reuse existing models and manage them through a proper governance process. If reuse is not feasible, new models can be developed, following guidelines for their creation and maintenance. However, data models are living documents that evolve over time, and governance of semantic standards is necessary to ensure these tools reliability and adaptability to the evolving needs of data space participants.

## Data Exchange

This building block focuses on the **transmission of data among participant of a data space**. The data exchange process involves a Transfer Process which progresses through a series of states, typically from "requested" to "terminated", aiming at managing transitions between providers and consumers in a systematic and transparent way.

To foster accurate communication and overcome technical interoperability barriers, data exchanges are based on clear protocols, which firstly regulate transmissions inside the data space, but through some of their principles can be extended to operations between different ecosystems. In a data space:

- Application Program Interfaces (APIs) or other connection mechanisms must be defined and published in the data asset catalogue, referencing the semantic meaning of the exchanged data to ensure its complete understanding by participants.
- Data exchange can be classified as push or pull transfers, depending on whether the provider initiates the data transmission (push), or the consumer is the first retrieving data. In both cases, the transfer duration and management are affected by the type of data exchanged, which can be finite (fixed datasets) or non-finite (continuous streams).

- It is also crucial to choose the proper transmission method based on industry standards. Furthermore, a preferred group of transfers should be identified to enable a quick connection between adhered data spaces.

The adoption of a common protocol, including the syntax and the sequence of the interaction, enables data exchange between participants within the same data space or in federated data spaces through the definition of APIs. As a general approach, a data exchange protocol must support the following functionalities, some mandatory while others recommended:

- **Efficient data transmission** (mandatory): data exchange begins after an event or a user's request. Identification and authorization are needed to enable the transmission.
- **List of capabilities for data asset operation** (mandatory): the data space governance approves data exchange protocols, which must be published in the value creation services and made available within the data space and in other ecosystems. Moreover, for data assets which may eventually require it, older versions of the exchange protocols should be available.
- **Querying of complex data** (recommended): a query language or a similar mechanism is necessary to support complex requests across various data structures.
- **Data retrieval endpoints** (recommended): the choice of a protocol depends on the usage scenario. Specialized mechanisms should ensure data integrity and efficient transport when dealing with large volumes of data.
- **Triggered exchange mechanisms** (recommended): alerts should be implemented to notify participants of data updates or modifications, enabling the right data provision without the need of user's requests.
- **Information retrieval in federation scenarios** (recommended): in federated ecosystems, specific exchange protocols of each data space must be available. A list of accepted protocols and versions should be shared for data transmission across spaces, possibly with a dedicated data model for quicker protocol negotiation.

## *Provenance and Traceability*

This building block offers guidelines for **supporting observability, provenance and traceability within the data exchange**. It may be necessary to track transactions occurring within the data space or identify who has accessed certain data. Therefore, some use cases – especially those involving high-value data – require additional metadata alongside the shared data for auditing and compliance purposes. In a data exchange, two phases may be distinguished: the control phase, involving transactions related to data-sharing contracts, and the actual data-sharing phase, where data is properly transmitted. While observability refers to the ability to **monitor and manage contracts**, and provenance tracking concerns the **sharing and usage of data**, it is important to notice that each data space participant is responsible for them both, thus they are not only technical features, but are strongly connected to regulatory and contractual compliance. For this reason, provenance and traceability standards and guidelines are spreading. Moreover, they are crucial elements for increasing trust among participants in a data space. Since data can be sensitive and reveal valuable information, it is indeed essential that the third party receiving it, due to observability, is also trustworthy.

A more detailed classification may be provided as follows:

- **Provenance tracking (data lineage)** refers to the backward-looking direction of a data value chain, e.g. the origin and processing of the data received by a consumer. This tracking includes:
  - ***Ownership***: the proof about ownership over a resource.

    - ***Custody***: the proof about owning a resource.
    - ***Location***: the proof of the resource origin or previous location.
    - ***Data Rights***: the proof that data handling, processing and sharing respect the rights of all stakeholders.
- **Traceability**, on the other hand, refers to the forward-looking direction of a data value chain, e.g. the usage made of the data which had been shared by a provider. Related information may be classified as:
    - ***Provenance traceability***, when tracking the origin of data.
    - ***Non-provenance traceability*** when the aim is not that of monitoring the provenance of data.
- **Observability** should be delimited, distinguishing activities which can be monitored within the data space from others outside of this feature goals but still important for trusted data sharing. In practice, observability includes:
    - ***Dataspace Protocol activities & states***: monitoring every state transition according to the data space protocol (e.g., data publication, contract negotiation, transfer process management).
    - ***Service telemetry***: monitoring additional data like service uptime and performance. As implementation-specific, each data space must agree on this topic.

Since each data product has specific requirements for provenance and traceability, typically in the form of metadata (i.e. about origins and quality assessment), they are creating new data and requiring a data model which, having to be interoperable by others, is subjected to rules and applications of any other data model, improving interoperability, access, and productization.

## Data Sovereignty and Trust

This pillar includes three building blocks, and its purpose is to **provide technical enablers with the capability to ensure the reliability and authenticity of the information shared, while also allowing participants to retain control over their data**. These goals are of utmost importance to foster is a crucial aspect of fostering trust between participants when interacting and conducting data transactions.

### *Identify & attestation management*

**Guidelines and technical mechanisms required to manage identities and other attestations within a data space** are the core of this block. It aims at ensuring that all participants can engage in the data space with confidence in its integrity, enabling them to **securely present, verify, store, and exchange attestations in a reliable and self-sovereign way**.

To this end, the Identify & attestation management building block provides the following capabilities to data spaces:

- **Machine-readable rulebook**: to enable automated compliance checks and interoperability across different ecosystems, the data space rulebook is provided in a structured and machine-readable format.
- **Compliance verification**: systematic verification of participants and services against the rulebook's requirements ensures coherence to the governance framework.
- **Trustworthy identities and attestations**: reliable identity and attestation mechanisms support secure onboarding and trusted exchanges.
- **Secure credential exchange and communication with wallets**: protocols for secure credentials sharing while maintaining data control are established.
- **Adoption of recognized standards**: consistency and interoperability are also fostered by alignment of this building block to widely accepted technical standards.

Within a data space, there are three main types of attestations for participants:

- **Identity attestations**: they involve a unique identifier associated to one or more attributes that uniquely describe an entity in a specific context. A participant may have multiple identities, but each of them is linked to a distinct identifier. Credential verification methods and compliance with standards contribute to identity trustworthiness.
- **Membership attestations**: after validation and compliance verification checks, the governance authority issues a membership credential, which confirms successful onboarding. This can also facilitate federation between data spaces when participants are recognised ad valid members according to the same rules.
- **Compliance attestations**: these attestations confirm adherence to sector-specific regulations and criteria defined by the governance framework.

Identity and attestation management in a data space ensures that organizations and individuals are properly identified and authorized through legal identity verification, personal identity checks, and sector-specific attestations. These requirements ensure trust and compliance within the data space. When defining the data space rulebook, the governance authority thus establishes a **conformity assessment scheme, whose outcomes are attestations of conformity to the designed rules and procedures**. Furthermore, credential schemas that facilitate secure access to the assessment results are also defined. The rulebook sets up also rules for credential validity and revocation. Moreover, a data space registry stores the lists of accredited trust anchors, trust service providers, trusted data sources, and notaries, along with the schemas for data space credentials. A credential store service, on the other hand, facilitates the management of verifiable credentials and other key information, allowing participants to present credentials to other data space participants and to validate credentials received from them. **The credential holder keeps exclusive control over the recipient of their verifiable credentials**, as well as the conditions under which they are exchanged.

## *Trust Framework*

Trust is a prerequisite for the most critical processes of a data space (e.g. identity verification, attestation validation, and data management according to the rulebook). This building block defines the key components of an effective trust framework, through which enhancing security, accelerating trust decisions, and supporting data exchange, thus fostering the growth of the data space.

The general elements of a trust framework are the policies, rules, and standards outlined in the rulebook, along with the procedures for collecting, organising and verifying information to support trust decisions. Moreover, it is necessary to define both a set of criteria for participants, linked to regulatory, business and technical requirements detailed in the rulebook, and some processes and technical mechanisms for their validation. Trust elements should provide to a data space the following capabilities:

- **Validation and verification of participants and services** to guarantee that all participants and services adhere to the established rules, promoting trust among participants and ensuring reliable and transparent data transactions.
- **Governance enforcement** of conditions and rules set by the data space governance authority to increase transparency and interoperability.
- **Definition of trust entities** providing them with clear guidelines.
- **Data space registry integration** offering mechanisms to store the data space rulebook, schemas, and the list of accredited trust entities.

For this purpose, data space participants may play some of the listed roles:

- Authoritative entities (e.g. governmental bodies) for which trust is assumed are referred to as **trust anchors**. Their trustworthiness is accepted by data space governance authority in relation to a specific scope of attestation.
- Legal or natural persons deriving their trust form one or more trust anchors are **trust service providers** (TSPs). They are chosen by the governance authority as parties eligible to issue attestations about specific objects, thus they can provide and preserve thematic certificates.
- The governance authority may accredit one or more **notaries** when a designated TSP is unable to issue cryptographic material and allow them to validate claims according to objective evidence form trusted data sources.

It should be noted that existing generic trust frameworks, such as those provided by IDSA Rulebook, Gaia-X, iSHARE or Ayra can serve as foundational models. Data spaces can thus leverage or change them introducing more stringent requirements and implementing additional rules into the rulebook.

## *Access & Usage Policies Enforcement*

The purpose of this building block is to **establish core policies that govern data management in data spaces**. They mainly relate to **access control**, determining who can access data and what framework establishes authorization conditions, **usage control**, specifying what action can or cannot be performed on data and how usage is controlled, and **consent management**, establishing processes and workflows between data rights holders and data subject for consent verification and revocation.

Key capabilities for an effective access and usage policy enforcement by the data space participants include:

- **Policy transformation and implementation**, which ensures policies are converted into machine-readable formats supporting compliance checks and enabling interoperability.
- **Policy management and governance**, facilitating the creation, update, and enforcement of policies concerning access controls, usage parameters and consent protocols in line with the governance framework.
- **Compliance tracking** to ensure adherence to established policies and audit data processing activities.

Data governance policies synergically cooperate in the workflow of a data transaction from initial request to final data access, but they operate in two distinct phases, each focusing on specific aspect of the policy lifecycle:

- In the **negotiation phase**, a data receiver has just identified a data product which is relevant to their objectives. A machine-readable technical agreement between the data provider and receiver is to be stipulated. The receiver's data transfer request goes through various architectural components and is processed based on policy evaluation and enforcement. In case of validation, a contract is issued, and data can be processed, thus completing the transaction.
- In the **enforcement phase**, the whole data transaction execution is concerned. The purpose of this phase is to evaluate the relevant policy rules to determine if the transaction can proceed (eventually verifying the validity of the stipulated contract). Diverse types of rules are evaluated at different process stages:
    - ***Access rules*** are primarily checked by the data provider before any exchange. The recipient may also verify them when receiving data.
    - ***Usage rules*** are assessed by the recipient during data usage, and they may require ongoing evaluation throughout the usage lifecycle.
    - ***Consent rules*** involve both provider and recipient in evaluating third-party consent, since it might be revoked during the transaction.

To guarantee effective policy enforcement and seamless operation in the data space, participants must implement the necessary access control and policy mechanisms. The governance authority facilitates the implementation of compliant, interoperable, and secure policies, reducing effort and ensuring adherence to governance standards through:

- **Policy templates and validation tools** to ensure policy correctness according to governance requirements.
- **Policy engine integration support**, recommending effective technologies and providing reference implementations.
- **Architectural guidelines** for integrating connecting participant systems and managing consent and data access.
- **Verification and compliance tools** to ensure policy enforcement and detect non-compliance.

## Data Value Creation Enablers

The last pillar aims at providing the **technical enablers to generate value among the data space participants through data sharing**. The following pages will focus on the three building blocks that facilitate the awareness of data, services and offerings available in a data space among its participants.

### *Data, Services and Offering Descriptions*

This building block aims at **outlining specifications, tools and technologies for metadata of data products and services**. In order to create an effective data sharing ecosystem, the description of data, services and offerings is created using **high-quality machine-readable metadata**, accessible both to human participants and to software. For ensuring discoverability and interoperability of these objects, metadata should meet criteria such as:

- **Application of FAIR principles**: by adhering to findability, accessibility, interoperability and reusability criteria, as well as to compliance, metadata and their related data products and services can be easily exchanged and interpreted by various stakeholders.
- **Customer-centric approach**: data users should steer the creation and periodic review of metadata.
- **Standardisation**: by following widely accepted standards, metadata enable unified search and discovery experiences.
- **Automation and quality control**: to reduce human mistakes and validate accuracy and consistency, automated metadata generation and quality control are recommended.

As already mentioned, metadata may describe diverse kinds of objects in a data space:

- **Data products**: a data product is made of datasets and data services. The description of datasets should capture all their features (e.g. format, structure, technical requirements, series in which they may be organised). On the other hand, services allow operations for selecting, extracting, combining and transforming those collections of data, thus their description ensures that a service can be discovered and used in the data space, and outlines how users can operate on datasets and what kind of output they should expect.
- **Technical services**: this category includes participant agent services, federated services and value creation services (e.g. by data quality enhancement, data analytics, data visualisation). Similarly to data services, metadata of value creation services' description cover general information, governance and compliance requirements, lifecycle management processes and relationships to other services or datasets.
- **Offerings**: an offering consists of data products and/or services together with their comprehensive description, including details about providers and creators, pricing model, licensing terms, rights and obligations. These descriptions are collected in catalogues which should be aligned with the technical and functional requirements of the expected users and consumers.

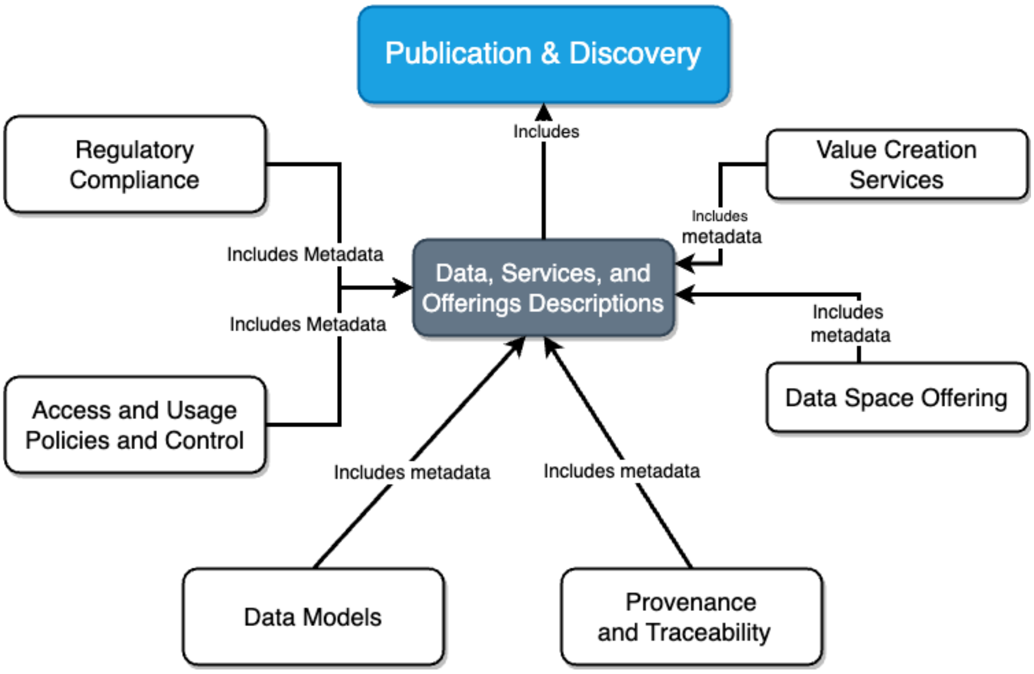



*Figure 19. The Interrelation between Data, Service, and Offerings Descriptions and other building blocks. Source: DSSC Blueprint v2.0*

For the creation, management and publishing metadata, data product providers typically use the DCAT (Data Catalog Vocabulary), which is designed to facilitate interoperability among data catalogues published on the web. They need to consider questions about **what the essential metadata attributes are, which policies should be included in those descriptions, and what kind of methodologies, tools and frameworks should be used for their updating**. Furthermore, providers may also define policies to regulate access, usage rights and other contract terms. These metadata enable data recipients to discover and evaluate the available products and services, through querying datasets, invoking data services, or downloading data in specific formats. The interaction between the parties is governed by established protocols and standards, ensuring interoperability and compliant data usage. Data spaces should thus support the following core activities:

- **Creating metadata**: to produce high-quality metadata, data spaces should provide necessary mechanisms to data providers, in compliance with the established standards, such as the DCAT and its extensions, supporting metadata alignment across European data platforms, ensuring metadata interoperability over regions, and facilitating standardisation in sectoral experiences. High-quality metadata provides information about spatial and temporal coverage, versioning, thematic features, packaging and compression format, technical specifications, endpoint descriptions and service URLs. Moreover, the Data Quality Vocabulary (DQV), an extension of the DCAT, provides a common approach to define quality attributes, including accuracy, completeness, consistency, provenance, update frequency, while the Open Digital Rights Language (ODRL) is used for metadata describing policies, terms of access and reuse, licensing conditions and restrictions. The sum of these features fosters trust and reliability between providers and consumers.

- **Validating metadata**: whether it is manual or automated through dedicated tools, a periodic validation process is conducted, in order to ensure that metadata aligns with established standards of DCAT and ODRL. Providers must ensure that descriptions are clear, unambiguous and fully accessible and understandable to consumers, regardless of their technical knowledge. Metadata can be considered validated after evaluating the consistency of definitions and terminology, checking clarity, accuracy, and standard compliance, and confirming metadata completeness.

- **Updating metadata**: since metadata can become outdated, incomplete or inaccurate, due to changes in the data, feedback from recipients, or new methodologies available, providers should continuously refine the descriptions of their products and services. The updating process develops in the following activities:
    - Reviewing metadata, by comparing current descriptions with latest data sources and standards, to identify discrepancies or outdated information.
    - Performing required operations, e.g. adding, deleting or modifying terms, examples and policies as needed.
    - Validating accuracy, by seeking feedback form recipients and users to validate the accuracy and completeness of metadata descriptions, as well as their effectiveness.
    - Ensuring compliance, remaining aligned to current legal and regulatory requirements
    - Adopting innovative technology or methodology in data products and services and reflecting them in their description, ensuring that users can leverage the latest tools and methods for their objectives.
- **Version control**: data spaces should be able to manage different versions of metadata in an effective manner. Version control allows providers to track changes with details about who, when, why and what kind of update was made, increasing transparency and accountability. The integrity of the descriptions is guaranteed through comparisons among different versions, which helps preventing both accidental or malicious changes. Moreover, storing previous versions of resource descriptions in an archive is essential for reference and audit purposes.

## *Publication and Discovery*

The descriptions of data, services and offerings presented in the previous building block are made available and discoverable within a data space through storage and publication in a catalogue. In full compliance with Article 33 of the European Data Act, the Publication and Discovery building block aims at **facilitating data transactions**, which are the core function of data spaces, **through descriptions of datasets and services, and terms and conditions under which they are offered**. On the one hand, it helps providers in managing the entire lifecycle of metadata and in exposing them as offerings towards data space participants as potential users; on the other hand, it supports consumers in identifying the best-matching offerings, according to their characteristics and terms of use, and request access to them.

To support the value chain of the offerings and enable smooth and dynamic transactions between providers and consumers, data spaces need to support the following functionalities:

- **Publication of an offering**: to interact within the ecosystem, providers must first register and authenticate themselves in the data space as participants. Whether authorised to publish in the offerings catalogue, they can select the offerings and publish them with their access rules. The catalogue informs the provider of the successful publication process, and the offerings can be thus discovered by potential consumers (all the participants of the data space or a subset of them, according to access rules, or even participants from other data spaces, whereas a global unique resolvable persistent identified (PID) is assigned to each offering).
- **Update of an offering**: already-published offerings can be modified by their owners, provided they are authorised to do so. Having updated the offerings and/or the access rules, the catalogue processes these changes, and potential consumers can discover datasets and services in their latest version.

- **Removal of an offering**: when allowed to modify contents in the data space, providers can select the offerings and/or the access conditions to be removed and the catalogue processes the removal. This is the end of the lifecycle of an offering. Consequently, the offering is no longer discoverable by other participants.
- **Discovery of an offering**: if correctly registered in the data space as a participant, a consumer can send a request to the catalogue, including the parameters to search the offerings. The catalogue then collects and sends relevant objectives accessible to the potential user. When holding the PID, even an external agent can discover a specific offering; however, access to the data or service requires subsequent onboarding into the data space.

In a data space, each participant agent typically possesses its own catalogue containing its respective offerings. This configuration is referred to as "**decentralised catalogue**", and it is incumbent upon the consumer to query each catalogue individually. When the query is a catalogue request, the entire content of the catalogue is returned, whilst when it is a dataset request, a specific entry is given. To access a provider's list, however, a consumer must be aware of its existence and know its access endpoint, which is usually made available in the data space registry, as it collects all the ecosystem participants. It is important to note that data providers may impose restrictions on the visibility of their offerings. Consequently, not every consumer may discover the same results when querying a specific catalogue. Depending on its specific governance framework, a data space may also include a centralised publication and discovery component: this "**central catalogue**" would encompass the offerings available to all the participants and could be made accessible also to external actors, a strategy that may be conducive to foster the growth of the ecosystem. However, synchronisation mechanisms are to be implemented, to ensure coherence of the central catalogue with those of the participant.

## *Value Creation Services*

The following and last building block addresses the **technical aspects related to services designed to create value, in different ways and through different means, out of the data shared** in a data space ecosystem. These services have already been illustrated in the Data Space Offering building block from a business point of view, and in the Intermediaries and Operators building block, when dealing with their providers. From a technical point of view, value creation services can be described by DCAT, to make them easily discoverable and accessible. The definition of "Data Service" in the Data Catalog Vocabulary is somewhat limited when compared to the scope of these services, as it is described as a "collection of operations accessible through an interface (API) that provide access to one or more datasets or data processing functions". For this reason, some scholars argue that this kind of service does not completely adhere to the definition of a data space. However, the direct and inherited properties of this class can be used to describe many of the properties of value creation services. Since value creation in a data space may come from a wide range of different sources, these services can be classified according to their role, purpose, and specific functionalities in the overall data space. The taxonomy presented below follows these criteria.

| Category | Description | Examples of services included |
|---|---|---|
| **Core services** | As boosting the capabilities of the data space, they are typically transversal to use cases and thus the most common Value Creation Services | – Data visualization<br>– Data quality management, assessment and validation<br>– Technical enablers for automatic compliance<br>– Anonymization and pseudonymization |
| **Data handling services** | These services aim at facilitating access and use of datasets | – Data selection, extraction, combination and packaging<br>– Data processing and transformation<br>– Data delivery<br>– Data interpretation and reuse |
| **Value added services** | They add value on top of data products and transactions and facilitate a cost-efficient technical implementation of use cases | – Data fusion and enrichment<br>– Machine Learning models hosting<br>– Artificial Intelligence driven services<br>– Data ethics, fairness and transparency<br>– Customizable and on-demand services |
| **Infrastructure integration services** | These services enable the connection to external infrastructures through which data can be processed, stored and collected | – Infrastructure catalogue<br>– Infrastructure orchestration and load balancing<br>– Provisioning |
| **Application integration services** | They enable the connection to external applications deployed on top of data spaces, ensuring a smooth interaction between the ecosystems. This connection will respond to the needs of specific use cases | – AI integration services<br>– ERP/CRM integration<br>– Simulation environments<br>– Vertical industry solutions |
| **Business enablement services** | These services support business models in the data space (e.g. by providing automated payment, and/or by automating and securing legal agreements among the parties…) | – Billing<br>– Smart contracts<br>– Certifications |

Value creation often requires integration among different kinds of services to successfully address specific requirements of use cases. This process evolves from identifying the use case needs and selecting the services to be combined, to ensuring their interoperability, testing and optimizing their results. For this reason, **composition capability** is a key property for a data space as it leverages the modular nature of the proposed taxonomy to add new functionalities and adapt workflows thanks to synergies among services. A data marketplace is a typical example of service composition, responding to the needs of even several use cases based on sharing, trading and exploiting data products: a data space can provide a trustworthy ecosystem in which relationships between suppliers and consumers can spread and tools for negotiating data delivery and usage conditions and monitoring the exchange process are fully available.

Whether it is simple or composed, a value creation service might be properly presented by following an information model, which helps offering a description and guidelines for implementation of these services. The model includes:

- **General information of the service**, including a high-level description of the service, its purpose, its lifecycle management and maintenance, and specifications about governance and compliance.
- **Service architecture and data flow**, with a focus on technical components of the service and the management of data, from their importing to their harmonization.
- **Access and usage**, including usability, interfaces, and security control mechanisms.
- **Services composition**, focusing on dependencies with other services and integration points.

To effectively manage value creation services, a data space should thus include capabilities of:

- **General performance monitoring**, to track the results and impacts of operations associated with the use cases from both a functionality and a user-satisfaction point of view.
- **Services monitoring**, with tools, metrics, processes and standards for measuring the services' efficiency and enable risk detection and management.
- **Traceability, usage tracking and audit trails**, ensuring the recording and analysis of services execution and usage.
- **Services composition**, to define mechanisms and requirements for combination and/or composition of services, ensuring them interoperability and scalability.
- **Services orchestration**, by designing effective and balanced flows of activities, with the correct sequence of inputs and outputs.
- **Testing, prototyping and piloting**, by creating preliminary versions of the services to test and validate their functionalities, design and user experience, by gathering feedback, identifying potential issues and refining the developed solution before its final deployment.
- **Seamless flow of data between the data space and external applications**, guaranteeing a smooth and continuous transfer of information, while preserving integrity and consistency.
- **Security**, implementing measures to identify and manage vulnerabilities, preserving the integrity of services.

## From theory to practice

The architecture of the data space, as outlined in the preceding sections on the 17 Building Blocks of the Data Space Support Centre Blueprint, constitutes a foundational technical framework which enables a deeper understanding of the complexity and interconnectivity inherent in the data space model, while also supporting its broader adoption across the Alpine Region. By providing public and private stakeholders with a **shared vocabulary and conceptual foundation**, it underscores the relevance of **data spaces as a means to facilitate data sharing and value generation**.

Beyond the initial contextual analysis and explanation of operational dynamics, the study integrates empirical evidence through the collection of testimonies from professionals actively engaged in data space-related initiatives and use cases, primarily concerning Manufacturing and Water Management as strategical sectors for the economic growth and the well-being of EUSALP local communities. These insights contribute to the development of the study's third section - policy recommendations - by illuminating the perspectives and concrete experiences of diverse stakeholders within the data space ecosystem.

The interviews produced a **wide range of qualitative data**, offering **nuanced insights into the current state of data space implementation**. Findings indicate variation in the maturity of initiatives, from established use cases to preliminary or aspirational projects, and highlight a critical transition wherein many projects are shifting toward sustainable business models as public funding phases out. Engagements included participants from European lighthouse projects, enablers, and technology providers. These interactions revealed ongoing challenges in standardization, stakeholder alignment, and trust-building – essential elements for the scalability and long-term viability of data spaces. Establishing trusted environments for data exchange emerged as a particularly critical condition for attracting new stakeholders and fostering active data contribution.

The following pages present **9 case studies**, each of which has been analysed according to **13 criteria** which, on the basis of the theoretical framework outlined above, are particularly relevant in order to fully understand the characteristics of the experience or project described. The **form** used during the interviews to capture this content, later revised and integrated with information from reference websites, is included in the annexes at the end of this document.

Three projects (**COOPERANTS**, **Circular TwAIn**, **Eur3ka**) relate to the Manufacturing sector, including the as-a-Service approach; four relate to Water, of which two (**Water Data Space** and **WATERVERSE**) are more specific thematic data space initiatives, while the other two (**USAGE** and **AgriCS**) address this resource in the context of experiences primarily focused on sustainability. Finally, two use cases (**datahub.tirol**, **Truzzt Alliance**) are not related to these sectors, but are particularly interesting due to their high level of maturity and their location in the Alpine region - a geographical criterion that could not be taken into account in some other case studies, where the priority was given to the topic of interest.

# COOPERANTS

## Challenge

The COOPERANTS **(Collaborative Processes and Services for Aeronautics and Space)** lighthouse project addresses the challenge of digitising processes in the aerospace industry, with a particular focus on improving collaboration among diverse stakeholders in the space manufacturing sector, especially during the early phases of the product lifecycle. The project aims to develop more efficient and decentralised working methods and processes throughout the entire lifecycle of aerospace and space vehicles, thereby enhancing the competitiveness of this industry in Germany and across the European Union. Key challenges identified include data exchange between stakeholders, protection of intellectual property, compliance with regulations (such as export control), and the complexity of the supply chain. This initiative is part of the broader effort by the German Federal Ministry for Economic Affairs and Climate Action (BMWK) to promote innovative and practical applications and data spaces within the Gaia-X digital ecosystem.

## Solution

To tackle these challenges, COOPERANTS is developing the digital platform "Digital Aeronautics and Space Collaboration Labs", a collaborative environment designed to enable secure, regulated data exchange among diverse stakeholders in the upstream aerospace sector. Built on the Gaia-X architecture, the data space supports interoperability, data sovereignty, and security, fostering effective collaboration between large enterprises, SMEs, startups, and research institutions. The platform was recognized in 2020 by members of the Bremen City of Space as a relevant Gaia-X use case within the Industry 4.0/SME domain. As part of its implementation, the project includes the development of a technical demonstrator to simulate real-world collaboration scenarios and test the benefits of this data-driven approach. The goal is to build an industrial ecosystem capable of working together more efficiently, unlocking new operational possibilities through enhanced data sharing.

## Development stage and use cases

The COOPERANTS project is currently in the implementation phase, with several use cases under development and testing. These use cases include the digitization of collaborative processes in the aerospace industry, the improvement of data sharing efficiency, and the development of new smart services. The goal of the project is to demonstrate the practical applications of the digital platform and its impact on reducing costs and risks, while enhancing collaboration and innovation. In this context, COOPERANTS is developing advanced digital services based on the GAIA-X data infrastructure, aimed at making production processes and collaboration within the aerospace sector smarter and more efficient. Among these services are agile manufacturing support systems that use digital twins to optimize production through real-time data integration, and collaborative platforms that simplify interaction between customers and suppliers. Furthermore, several pilot missions are underway to showcase the practical application of COOPERANTS technologies throughout the entire lifecycle of a spacecraft. Key examples include the collaborative design of spacecraft between concurrent engineering centres, the remote control of robotic systems through digital twins, and the optimization of computational fluid dynamics simulations to improve the digital design of aircraft.

## Geographical area of reference

The COOPERANTS consortium is composed of industrial, technological, and scientific actors active across Germany, involving both large enterprises and SMEs, as well as prominent research institutions. The project operates at both national and regional levels. Nationally, a support framework has been requested to ensure reliability and foster innovation within the aerospace industry. Regionally, COOPERANTS collaborates with specific territories across various German federal states to develop and test local use cases, enabling targeted implementation and fostering regional innovation ecosystems.

## *Partners*

The COOPERANTS consortium, coordinated by the German Aerospace Center – Space Systems (DLR-RY), brings together a strong network of industrial and technological actors from the aerospace manufacturing sector, including software providers and research institutions. The initiative spans several German regions: Bavaria, Baden-Württemberg, Bremen, Lower Saxony, North Rhine-Westphalia, Saxony, and Thuringia. Key industrial partners include Airbus Defence and Space, itemis, and OHB System, while SMEs such as neusta aerospace, RADIUSMEDIA, ScopeSET, and ZARM Technik contribute with their specialized expertise in components, machinery, and software development for the space sector. The startup Valispace adds innovative capabilities to the ecosystem. Scientific and research activities are supported by top institutions, including DLR, the German Research Center for Artificial Intelligence (DFKI), the Fraunhofer Institute for Machine Tools and Forming Technology (IWU), and the FZI Research Center for Information Technology.

## *Participants*

The project involves a diverse group of participants, including:

- Large enterprises, driving industrial-scale innovation and integration.
- SMEs specialized in the production of components, machinery, or software for the space sector.
- Private research and development entities, contributing with technological know-how and scientific support.

## *Data sharing model - architecture*

The COOPERANTS data space is built on the Gaia-X architecture, ensuring that the solutions developed are interoperable with other Gaia-X-compliant systems. This approach allows data to be shared and utilized efficiently across the ecosystem, facilitating large-scale collaboration between diverse stakeholders. In addition, the project adopts the Pontus-X technology stack, which enables the use of common data catalogues and further strengthens interoperability with other data spaces using the same technology. The main access interface to the data space is a marketplace, providing a centralized point for discovery and interaction.

## *Types of data*

The COOPERANTS data space hosts a wide variety of data from the aerospace industry, including both private and proprietary data related to industrial processes and manufacturing, as well as non-proprietary and public data used for research and development purposes. Among the data managed within the space are also sensitive datasets, particularly those involving intellectual property and regulatory compliance.

## *Interoperability*

Interoperability is a central focus for COOPERANTS, which adopts the Gaia-X architecture to ensure compatibility across different systems and platforms. The project promotes shared rules and the use of common technological standards to enable smooth integration and communication between stakeholders. COOPERANTS specifically addresses the challenges of semantic interoperability, aiming to improve the understanding and exchange of data across diverse actors and domains. By leveraging technologies such as the Pontus-X stack, the consortium ensures that its solutions are not only interoperable within the COOPERANTS ecosystem but also compatible with other Gaia-X-based and external data ecosystems. This approach helps overcome barriers related to data interpretation and standardized data exchange.

## *Trust*

To ensure trust and compliance within the COOPERANTS ecosystem, the consortium adopts rigorous governance measures for data management. Data security is a fundamental priority, and COOPERANTS is committed to adhering to all relevant European and international data protection regulations and compliance standards. Trust among participants is further strengthened by transparency in processes and a clear definition of data access and sharing rights. Within the data space, trust is also fostered through the adoption of technical mechanisms as outlined by Gaia-X, along with regulatory compliance and clear management of intellectual property.

## *Present and future sustainability*

The COOPERANTS project is currently funded at the European level, through the BMWK support program, with the aim of developing practical applications for the aerospace industry. The consortium is exploring various business models for the future, including subscriptions, pay-per-dataset, or transactional models where a portion of the fees would go to a central entity. Additionally, COOPERANTS is working on a growth plan to expand the number of participants throughout the year. The consortium is also focused on the future evolution of the data space, with a key emphasis on expanding participation and creating a sustainable business model to support the long-term growth of the project.

## *Obstacles*

The main obstacles encountered include the complexity of collaboration among different stakeholders, supply chain management, regulatory barriers, and the need to digitally integrate processes, especially for SMEs which often have limited resources and expertise in digital cooperation.

## *Policies*

To further support projects like COOPERANTS, it is essential that governments provide technical and financial support, promote the creation of shared standards, and facilitate the adoption of data space solutions. Additionally, public policies should focus on ensuring regulatory clarity (especially concerning intellectual property and export controls), supporting the digitalization of SMEs, and improving coordination between regional and national levels. Furthermore, the future of COOPERANTS project will depend on the success of initiatives aimed at promoting trust in data sharing while ensuring data security and regulatory compliance.

# Circular TwAIn

## Challenge

Circular TwAIn **(AI Platform for Integrated Sustainable and Circular Manufacturing)** aims to support the transitioning of the manufacturing industry towards more sustainable and resilient processes, as defined by the principles of Industry 5.0. On the one hand, this sector is actually facing challenges related to solutions lacking integration and interoperability, which hinder the adoption of circular and sustainable practises to reduce waste and optimise resource use. Moreover, fragmented and non-systematic implementations of AI technologies often obstacle end-to-end sustainability and slow down adaptation to new environmental criteria in the market. A holistic and trustworthy approach to sustainability must thus be promoted and hopefully adopted by all the manufacturing stakeholders.

## Solution

Circular TwAIn addresses these challenges by developing a novel AI platform for circular manufacturing value chains. By leveraging trustworthy AI techniques, interoperable circular Digital Twins are created, thus facilitating real-time monitoring and optimisation of manufacturing processes. Through its data space-based solution, the project facilitates secure and seamless data sharing and integration among a variegate ecosystem of manufacturers, recyclers, and technology providers, thus covering the entire product lifecycle. Circular TwAIn promotes a “circular by design” approach: from a technical point of view, by maximising the use and re-use of production waste, while from a cultural perspective, by providing tools and trainings for optimising decision-making, developing technological skills to properly manage innovation in products and processes, and creating new circular business models in the manufacturing sector.

## Development stage and use case

The project is currently in the last months of implementation phase and has a total duration of three years. Aiming to promote circularity and end-to-end sustainability through the integration of AI and Digital Twin in the data space approach, three industrial pilots – as three use case – have been developed.

- De- and Re-manufacturing of Li-Ion battery packs in e-mobility: this pilot addresses the lack of value chain optimisation in recovering functionalities and materials from automotive battery systems and aims to integrate innovative circular economy nodes, by collecting battery systems which have reached their end-of-life, assessing their reusability and eventually disassembling them to save units suitable for second-life batteries.
- De- and Re-manufacturing of consumer WEEE: to face the challenge of over 10 million tonnes of e-waste generated in Europe, with significant environmental damage and economic losses, this use case implements Extended Producer Responsibility (EPR), requiring manufacturers and importers of electronic components to take responsibility for the management of their products’ end-of-life.
- Petrol-Chemical Pilot: Hybrid Circular Twins for Process Industry: the complexity of production processes and the need for specialised equipment may hinder the integration of circular manufacturing practices. Through AI and Digital Twin technologies, this pilot focuses on the issue of sustainability by optimising resources and reducing waste.

## Geographical area of reference

The project development primarily involves the 11 nations represented in the partnership: Italy, Spain, Portugal, Germany, Switzerland, Luxembourg, Norway, Serbia, Cyprus, Greece and Türkiye.

Among these coutries, Italy, Spain and Türkiye hosted the three project's use cases. The issue addressed by Circular TwAIn, the scalability and replicability of its methodologies, and synergies with other EU-funded projects and initiatives expand the geographical area of reference to the broader European context.

## Partners

Led by Engineering – Ingegneria Informatica, the Circular TwAIn consortium involves technology providers, research organisations, sustainability domain experts and manufacturing companies. The partnership is composed of 21 members from 11 European and extra-Europe countries: Italy (Politecnico di Milano, GFT Italia, HAIKI COBAT, RaeeMan), Spain (Tecnalia, Revertia, AIMEN Technology Centre, Recyclia), Portugal (UNINOVA), Germany (Fraunhofer), Switzerland (SUPSI (University of Applied Sciences and Arts of Southern Switzerland), Switzerland Innovation Park Biel/Bienne), Norway (Sintef), Luxembourg (Netcompany-Intrasoft, Expert AI), Serbia (Nissatech Innovation Centre), Cyprus (Suite5), Greece (CORE Innovation Centre), and Türkiye (Teknopar, Socar).

## Participants

Both public and private entities participate in the solutions developed by the project. Research organisations and technology providers deployed the AI Application Modules and the Digital Twin Interface Modules, while companies provided data for the implementation of the three pilots in discrete manufacturing and in the process industry.

## Data sharing model – architecture

Circular TwAIn model consists of AI-powered Digital Twins consuming data within a data space based on the EU-funded OPEN DEI project principles for Data Spaces and their building blocks and on Simpl Data Space Design Principles, and able to support various use cases while storing historical information on components, products and processes. The architecture of the project is based on principles of modularity (developing software in modular and easily reusable structures), standardisation (as the adoption of existing framework and standards, also to ensure a consistent level of technical maturity), openness (predilecting open-source components), and hardware independence (exploiting scalability and elasticity typical of cloud resources). Circular TwAIn thus is composed of two different types of software modules that can be dynamically combined, extended, added and deployed according to the target application:

- AI application modules, containing both machine learning and deep learning functionalities, developed using open-source frameworks, and work as elementary building blocks of further solutions. These modules can either include pre-trained models to address common challenges in the industrial sectors or be customised and extended with additional training data.
- Interface modules: Digital Twins for Circular Manufacturing AI applications enhance AI application across the circular value chain. In order to train human-AI applications, relevant sources of data concern the manufacturing process and product, as well as the human operator.

## Types of data

The Circular TwAIn data space collects various types of manufacturing data concerning product lifecycle, industrial process (monitored by sensors and IoT systems), material composition for recycling and reuse optimisation, and environmental impact metrics such as energy consumption and gas emissions. This information feed both AI technologies and Digital Twin solutions to address the sustainability challenges faced by the project.

## *Interoperability*

Technical and semantical interoperability is fundamental for Circular TwAIn, which thus adopts existing frameworks to create a complex modular solution integrating different technologies of AI, Digital Twin and data space ensuring both conformity to standards and in force regulations and a proper level of technical maturity. To access the ecosystem manufacturing data should adopt specifical interoperable formats and protocols. Moreover, interoperability is enhanced through the integration of collaborative AI and seamless data sharing technologies.

## *Trust*

Circular TwAIN is based on the concepts of Collaborative AI and Seamless Data Sharing, both closely related to the theme of trustworthiness, which can make AI-based manufacturing processes more attainable to humans and strengthen relationships between all actors and thus the circular manufacturing system as a whole. To address the challenges related to this issue, the project emphasised data sovereignty, security and privacy measures for both the data space and AI tools. Stakeholders and participants can therefore rely on the integrity and confidentiality of the data shared within the platform. Furthermore, the three use cases developed in the project are aimed at increasing trust in the benefits of this ecosystem and of integrating sustainability criteria and emerging technologies in manufacturing processes. Within the project, a diverse range of training courses has been made available in order to support the upskilling of manufacturing workers with the essential knowledge and capabilities to manage innovative production processes.

## *Present and future sustainability*

Circular TwAIn is a 36-month project funded by the European Union's Horizon Europe Digital, Industry and Space programme. Since the end is scheduled in July 2025, the consortium is currently working on an exploitation plan to ensure continuity to the developed platform and engage new test areas – the spread of Digital Product Passports, for example, may ensure great interest in the project approach. However, to this aim, synergies with other initiatives have already been explored. Circular TwAIn is member of AI4SAM, together with Horizon Europe projects s-X-AIPI and AIDEAS. The consortium shares a common focus on using AI as a transformational tool to support circular production in the manufacturing and process industries and expand the positive impact of these initiatives beyond the project level, boosting awareness of the most significant research outcomes. Moreover, Circular TwAIn supports the Didactic Factories by directly managing four of them. These factories provide hands-on training, access to technical expertise, and test-before-invest services, with a focus on Industry 4.0, digital transformation, circularity and AI. Future sustainability may also be fostered by Circular TwAIn Marketplace, where are published the solutions leveraged by the project partners and a better overview of benefits related to Digital Twins and AI algorithms applied to manufacturing products and at production level.

## *Obstacles*

Since promoting the adoption of various emerging technologies, the project encountered several significant obstacles. The integration of AI technologies into existing manufacturing processes in order to foster circularity and sustainability required adaptation and customisation to the needs of the final users. On the other hand, data needed to be standardised across participants and stakeholders to support sharing practises: thanks to the consortium competences, Circular TwAIn data space could adopt common European interoperability standards and include strong security measures aimed at fostering the trustworthiness and transparency of the ecosystem. A change in the mindset, however, was the main obstacle to the successful project implementation: the pilot actions were addressed to this major scope, as well as the availability of resources and results in the marketplace section and the deployment of training and upskilling materials.

## *Policies*

The Circular TwAIn project has shown the importance of fostering the integration of technologies such as AI and Digital Twin in implementing circular manufacturing process chains and waste optimisation. Policymakers should thus support the integration of data space solutions to enable the creation of ecosystems where the integration of emerging technologies and the willingness to share data promote the logic of “as-a-Service” and the principles of sustainability. Incentivising this kind of experiences also between SMEs may promote a wider adoption of this approach, thus concretely facilitating the dual transition to digital and sustainable practices.

# Eur3ka

## *Challenge*

Eur3ka **(European Vital Medical Supplies and Equipment Resilient and Reliable Repurposing Manufacturing as a Service NetworK for Fast Pandemic Reaction)** aimed to respond to the severe deficiencies in the production and supply capacities of medical equipment emerged during the COVID-19 pandemic. From the manufacturing sector's perspective, the main challenge related to such an extraordinary event concerned the need for a more agile and coordinated approach to help traditional systems in coping with rapid changes in demand of products, avoiding shortages and dangerous inefficiencies. Eur3ka therefore focused on developing a trustworthy framework as a unified response sytem that could quickly adapt to COVID crisis, guaranteeing the constant availability of medical supplies, and to eventual other disruptive events.

## *Solution*

A multifaceted solution to these challenges was proposed by Eur3ka. To foster effective and coordinated response to sudden increase in demand of medical equipment, the project consortium developed a trusted European Manufacturing Repurposing Network, including the creation of a Coordination Platform and a Plug & Respond (P&R) Network as a data space ecosystem through which facilitating rapid repurposing of manufacturing capabilities. Seamless connection between all kinds of manufacturing stakeholders was fostered by the integration of Industry 4.0 standards, open automation, and innovative digital technologies. To ensure trustworthiness to the solution, international standards and technical components were adopted, and a common governance model to promote collaboration among partners and participants, which is essential during crises, was developed.

## *Development stage and use case*

The project started in December 2020 and ended in February 2023. After a first consultation of domain experts and the review of existing components from the partners, the consortium developed the Plug & Respond (P&R) repurposing resource coordination framework and validated it through various tests, which may considered as the project use cases, concerning, for example, the production of surgical masks and diagnostic technologies.

## *Geographical area of reference*

Europe was the main geographical focus of the project, which aimed to enhance the continent's ability to respond to medical supply crises, as happened during the pandemic. Strong international cooperation ensured widespread benefits and foster the global scalability and replicability of the solutions and methodologies developed.

## *Partners*

Involving 25 partners, Eur3ka consortium was led by Engineering – Ingegneria Informatica and equipped with a wide range of competences and expertise needed to effectively address the challenges of the project. 12 European and extra-Europe countries were represented: Italy (Seacsub, STAM, Politecnico di Milano), Spain (ATOS IT solutions and services Iberia, Software Quality Systems), Portugal (Unparallel Innovation), Germany (ATOS Information Technology, Digital Hub Management, Fraunhofer, International Data Spaces), Switzerland (Eidgenoessische Technische Hochschule Zuerich), The Netherlands (Brainport Industries Cooperatie), Denmark (SVM Automatik, Teknologisk Institut), Norway (universitetet i Oslo), Finland (Visual Components), Luxembourg (Netcompany-Intrasoft), Romania (Siemens srl) and Israel (Siemens Industry Software).

## Participants

The project participants were both public and private entities. Research organisations and technology providers were responsible for the PoC and the digital tools which connected re-converted providers and consumers of health products. On the other hand, manufacturing companies across Europe were involved as project use cases and benefitted from data sharing practices.

## Data sharing model – architecture

The architecture of Eur3ka consisted of a federated digital infrastructure enabling efficient and secure data sharing across European manufacturing networks. Since dealing mainly with proprietary data, robust protection measures and standards were integrated in the model, as well as Industry 4.0 technologies, including Digital Twins and International Data Space connectors to guarantee interoperability within the ecosystem.

## Types of data

Eur3ka data space predominantly collected information from manufacturing companies. Data thus concerned the production capacity, the present and programmed workload, the manufacturing capabilities (essential to evaluate the possibility to convert the previous production according to health crises), and the supply chain. The sensitivity of this information required the consortium to design a strong trustworthy technological infrastructure.

## Interoperability

In its consortium, Eur3ka involved members of the community which is driving the development of data space model, thus ensuring that the criterion of interoperability was properly addressed during the project implementation. For this reason, both standardized modular manufacturing reference architectures and IDSA tools were adopted. Fostering interoperability ensured easier coordination among companies' networks in response to a sudden increase in demand.

## Trust

Addressing the issue of trust in Eur3ka meant designing a comprehensive solution where technology and governance frameworks were aligned in supporting the creation of a secure ecosystem. By adopting international standards and trustworthy technical components and by involving the Digital Factory Alliance, the project thus emphasised data sovereignty and the protection of sensitive information from manufacturing companies. The Plug & Respond Network developed by the project thus succeeded in efficiently connecting providers and consumers of health products, allowing the demand for goods during crises to be easily met.

## Present and future sustainability

Until 2023, the project was funded by the European Union's Horizon Europe Industrial Leadership programme. Since it tried to address the contingent health crisis of Covid pandemic, it is no longer operational. However, its innovative approach to repurposing manufacturing capabilities and enhancing supply chain resilience has proved to be successful, thus it may be globally scaled and restored in case of future crises, ensuring a faster response to emergencies.

## *Obstacles*

From a technical point of view, the project had to face some difficulties concerning rapidly repurposing manufacturing processes to meet medical supply demands within a manufacturing as a service framework. Another important challenge was related to data security and trust in the ecosystem. These challenges could be overcome thanks to the availability of a wide range of competences and expertise in Eur3ka consortium, a robust internal collaboration and adherence to trustworthy standards, and the successful implementation of the P&R network and use cases.

## *Policies*

The effectiveness of such initiatives depends on a strong collaboration across sectors and countries: by promoting standardisation and interoperability, data spaces can enhance global response to crises. Moreover, Eur3ka has addressed some final policies concerning both performance dimension and people dimension, in the short term, long term or as rapid actions. To strengthen an ecosystemic approach, recommendations support establishment of partnerships, diversification of supply chains, investments in technologically innovative infrastructures, and implementation of upskilling programmes.

# Water Data Space

## Challenge

The **Water Data Space** was a Danish national initiative to address the macro-challenge of managing water data more effectively and demonstrate the benefits of better data sharing, with the objective of creating the world's most sustainable water cycle in Denmark. To this aim, water sector needs to value data as a crucial resource to increase efficiency, reduce waste, limit the use of environmentally hazardous substances, and adapt to climate change and its extreme events. However, a lack of governance, business and data models, legal and technical frameworks for data reuse, and limited availability and accessibility to data and best practices for data sharing, make it difficult to fully exploit data. To raise awareness of this issue, the project identified six societal challenges related to water data, ranging from weather balance to climate, from energy to circularity, and from health to investment in infrastructure. As all these challenges involve data from both locked ecosystems at water utilities and industry and from the surrounding environment, the Water Data Space had to find a way for external stakeholders to access relevant data in closed ecosystems. Moreover, working with water data means dealing with fragmentation across utilities in terms of digital skills, priorities and trust in sharing practices as well as a lack of both common data standardisation and common rules / infrastructures to govern data exchanges.

## Solution

The solution developed by the project consisted in the creation of an open digital ecosystem with a strong collaboration and governance model to support innovation in the water sector by facilitating the collection and processing of data from different sources. In line with the European Commission's advice to promote green and digital transformation for the benefit of all stakeholders, the Danish Water Data Space thus created a trusted framework where multiple data sources could be accessed and shared for multiple uses, both public and private, with the aim of developing a sustainable and efficient national water cycle. This solution benefited from the high level of digitalisation in Denmark and from the possibility of free access to environmental water data. However, in order to demonstrate the value and concrete potential of the data space approach, four use cases were developed, addressing priorities and needs by tailoring the basic framework to meet the specific requirement of the involved stakeholders. The Water Data Space attempted to foster change in water management by focusing on what hindered innovation in the sector, but it was only a step towards a comprehensive solution to the challenges previously described, which would require accessibility to water data across the supply (and value) chain, leveraging of political and regulatory focus on Environmental Health and Safety (EHS) to achieve greater data harmonisation, and promoting ever new pilot projects to increase the potential of data sharing.

## Development stage and use case

The project had a total duration of three years and resulted in the creation of an open data space between 2021 and 2023, enabling the water industry to innovate and streamline based on data. This main objective was declined in four different use cases, which focused on:

- Automated reporting: optimising the workflow for annual reporting of benchmark utility data to authorities was the goal of this use case, which faced a lack of standardisation, time consumption and the presence of errors in the reporting process. Automating this process showed great potential for enhancing efficiency within water utilities.
- Detention pond analytics: digital technologies were there used to provide an online overview of the pond performance, improving environmental conditions in creeks and streams. Real-

time data and insights supported better management of detention ponds, ensuring their effectiveness in mitigating flooding and maintaining water quality.

- Environmental Health and Safety (EHS) source tracking: this use case focused on collecting and visualising data from water dischargers to identify potential emitters of harmful substances, helping authorities and stakeholders to more effectively monitor and managing these risks.
- Water quality prediction: analysis of historical and real-time data enabled accurate predictions of water quality in various water bodies. This solution supported proactive water management and pollution control.

## Geographical area of reference

The Water Data Space was led by Water Valley Denmark, an alliance serving as a national unifying initiative driving the development and innovation of solutions for the water sector. The geographical scope was therefore all of Denmark, and the four use cases developed in the project were not conceived as local pilots, but addressed national challenges related to water management.

## Partners

The project consortium consisted of 11 Danish partners who brought competences to effectively address the challenge of shaping an innovative national water data space. Water Valley Denmark, Systematic, Aarhus Vand, Danmarks Tekniske Universitet, DHI Danmark, Danmarks Miljøportal joined a steering committee of organizations with expertise in data management, digital technologies and water resource management. This group worked together with research centres (Aarhus University and Aalborg University) and national water utilities (Herning Vand, HOFOR, and VandCenterSyd). The involvement of different stakeholders was crucial to ensure a comprehensive approach and successful development of the project.

## Participants

The Water Data Space involved both public and private entities, e.g. utilities, research organizations, large enterprises and SMEs. Participants contributed with datasets and data products related to water management, as well as with technological solutions and services to address the specific challenges of the use cases. Security providers were involved in the technical design of the platform. The project paid particular attention to the difficulty of incorporating SMEs into the data space, due to the predominant presence of public authorities, utilities and large consultancies. Additional efforts were thus made to include these participants, as small and medium-sized enterprises are essential to contribute to economic growth and sectoral innovation.

## Data sharing model – architecture

The Danish Water Data Space developed an initial model inspired by the general design principles of the European Data Strategy (2020) and created a digital framework where multiple data sources can be accessed and shared for multiple uses, thus with a focus on interoperability, scalability and reusability. Smooth interaction with the data available at the Danish Environmental Portal was a primary goal of the architecture, which therefore included specific tools and solutions to ensure security and privacy. In line with the existing technology standards, the framework allows real-time data sharing and analytics, as well as accessibility through APIs. As the technical backbone of the Water Data Space, the hard infrastructure orchestrates and enforces the data sharing agreements, while the soft infrastructure ensures the effectiveness of the architecture, by defining a streamlined process for integrating and accessing data assets.

## *Types of data*

The Water Data Space hosted data repositories from companies, databases from utilities, open data from public authorities, and research data from universities. The diversity of participants in the ecosystem ensured the availability of a wide range of information related to three main domains: water in aquatic ecosystems (with most data produced by public monitoring), water in pipes (utilities, industry), and water in cities (often related to cloudbursts). These data have different characteristics in terms of data points and time series, depending on wheter they represent assets, operations or economies (e.g. geographic data are many data points, few changes, while sensors may have few data points, values changing every second). Effective management of water data is essential to improve decision making.

## *Interoperability*

Since water is often linked to other sectors (e.g. energy, biodiversity, urban planning, food, etc.), data sharing can involve many different stakeholders and administrative structures, thus increasing the need to harmonise data for interoperability. The digital ecosystem implemented in the project addressed this challenge integrating data management tools and methodologies to seamlessly combine data from different sources and use them together. The participation in the consortium of organisations with expertise in digital technologies was essential to promote and assess the FAIRness (Findable, Accessible, Interoperable, Reusable) of water data. Through its use cases, the project drew attention to three issues related to interoperability that need to be addressed as enablers of a broader digitalisation of the sector:

- Easy-to-integrate and free data standards should be more widely developed in order to make data usable across actors and sectors and to prevent incorrect interpretations and conclusions.
- The spread of standard API protocols for data exchange would facilitate the integration of different data sources and lower the barrier for smaller organisations wishing to start sharing their data.
- Shareability should become a fundamental criterion when choosing the format for data collection.

## *Trust*

The issue of interoperability was closely linked to the issue of trust: the lack of trust between data providers and data users is recognised as one of the most critical reasons for the suboptimal use of data, and it hinders any efforts to increase shareability and interoperability, consequently limiting availability and accessibility of data. The use cases implemented aimed to build convincing business cases and to demonstrate the benefits of collaboration between different entities to solve common problems. These initiatives specifically targeted water utilities, which were hesitant in sharing their data due to a lack of confidence in how it would be used. The project focused on data security and privacy integrating high levels of protection and blockchain technology to provide traceability and transparency in the ecosystem. In addition to the technological aspects, trust was also fostered by establishing a strong and efficient governance model.

## *Present and future sustainability*

The Water Data Space was supported by the European Regional Fund REACT and the Danish Board of Business Development. As the project was successful in demonstrating that data sharing can create value for the whole water ecosystem, it continues today in the Lighthouse Water-tech domain with Water Valley Denmark association supporting new pilot projects both to illustrate the potential of data sharing also in other water-related sectors and to overcome barriers and challenges. With new funding from the EU and the Danish Board of Business Development, best practices are now being explored, and further learning and trust is shared between stakeholders. Building on the previous project, the Water Data Space Lab will create new digital business opportunities for water technology SMEs by strengthening data platforms and leveraging open data. Until February 2026, companies are invited to develop innovative solutions tha address critical issues and needs in the water industry and to develop a prototype of a water data sharing architecture, in order to concretely implement a water data space. The sustainability of the solution will therefore depend on many factors relating to technology, business and governance. In fact, the consortium is discussing about who, among the various stakeholders (public authorities, private organisations, and water utilities), should lead the process of implementing the data space and thus its ecosystem.

## *Obstacles*

As the implementation of a data space is both a mindset shift and a change management process, one of the obstacles encountered during the project was to foster commitment to working processes and the necessary investments in the partnership. A collaborative stakeholder analysis was carried out, which led to better understanding dynamics that foster or hinder data sharing practises. Lack of trust - mainly due to a lack of successful examples and best practices - emerged as the most critical issue: the development of the four use cases highlighted that the value creation of data sharing cannot be fully understood before data sharing starts. At the same time, it is difficult to develop technology solutions broadly applicable in the water sector, as digitalisation is slowed by fragmentation among the stakeholders and a lack of standardisation and of quality of data from the utilities. Through their involvement in the project, participants have been involved in creating a common vocabulary and discussing standards for data collection and sharing standards, thereby supporting digital literacy among water stakeholders. Another important obstacle faced by the project was the involvement of the SMEs, especially as providers of innovative solutions. In fact, they typically lacked an overview of where to find data, wasted a lot of time waiting for data from utilities and industry, and had to deal with non-standardised data. Actually, helping SMEs enter this ecosystem is one of the main objectives of the Water Data Space Lab.

## *Policies*

The Water Data Space project highlighted that data sharing is still limited, due to trust issues, technical obstacles, and lack of appropriate regulations to govern these exchanges. For this reason, the consortium defined seven policy recommendation, concerning the needs to:

- Define common data standardisation, at national or European level.
- Develop regulations and infrastructure for data sharing.
- Identify intermediaries with IT and governance capabilities to manage the Water Data Space.
- Promote a vision of water as a strategic sector in the EU digitalisation strategy.
- Organise the Water Data Space through a decentralised infrastructure.
- Regulate the ecosystem in order to enable FAIRness across the value chain.
- Support first-mover cooperation via innovation projects and partnerships where data spaces are conceived as a tool for trusted and secure exchanges.

# WATERVERSE

## Challenge

**WATERVERSE (Water Data Management Ecosystem for Water Data Spaces)** addresses the challenges of data fragmentation, lack of interoperability and high costs associated with data management in the water sector across Europe. It is primarily aimed at water utilities, which are typically characterised by a conservative approach to data sharing, fear of losing control over data and limited interest in investing in improving data quality. For this reason, the project aims both to overcome the technical barriers to effective data valorisation and to support a mindset shift towards viewing data as a resource. The challenges therefore relate to accessibility, affordability, fairness, usability and interoperability of water sector data. These issues can be seen as challenges in identifying and integrating tools and methodologies to ensure the effectiveness, efficiency and security of data sharing practices, while promoting their scalability, replicability and business applicability.

## Solution

WATERVERSE addresses these challenges by developing a Water Data Management Ecosystem (WDME) that facilitates data management practices to improve the resilience of the water sector, while increasing the perceived value of data by water utilities. The project aims to develop this solution in four steps, starting with the characterisation of the overall water data management context by mapping key stakeholders and needs, and addressing issues related to the quality and architectural design of the data management system. The WDME will then be built as a modular solution of tailored tools and resources, integrating cybersecurity measures and promoting interoperability. Fairness metrics will then be developed to assess the quality of both the data and the ecosystem, as a prerequisite for a future implementation of a water data space. Finally, six pilot projects will be implemented across Europe, each addressing a specific local challenge related to water. Thanks to this strategy, the impact of the project is expected to reduce health risks, water losses and emergency response times, improve operational efficiency and increase the availability of open access datasets.

## Development stage and use case

The project is currently in the last months of implementation phase and has a total duration of three years. WATERVERSE was not aimed to create a data space, but to foster a Water Data Management Ecosystem with tools for data collection and quality improvement, as a precondition for future sharing and valorisation. To spread this mind shift, six pilots – as six use cases – have been led across project partners' countries, each of them focusing on a specific challenge, from predicting water quality and its impact in the treatment steps in The Netherlands, to digital village twin for flood protection and territorial management in Germany, from leveraging the potential of water digital twin and water analytics tools in Cyprus, to addressing the challenges of Combined Sewer Overflow performance in the United Kingdom, from managing and innovating the integral water cycle in Spain, to developing smart water tools for efficiency enhancement in Finland.

## Geographical area of reference

The pilot tests of the WATERVERSE project have been implemented across six European countries, and each of them addresses a specific local water management challenge. Pilots are hosted in the District Local Government Organisation of Limassol (Cyprus), in the region of Andalusia (Spain), in the village of Etteln in the municipality of Borchen (Germany), across Finnish territory, with a focus in the cities of Espoo and Helsinki (Finland), in North Holland region (The Netherlands), and across Devon and Cornwall (United Kingdom).

## Partners

The project consortium involves 17 partners, bringing a wide range of competences and expertise to effectively address WATERVERSE challenges with a multidisciplinary approach. Research organizations, water utilities, technology providers in the water domain, and innovation companies are included in the partnership. 10 European countries are represented: Greece (CERTH - Centre for Research and Technology – Hellas), Spain (Cetaqua Barcelona, Eurecat Technology Centre, Hidralia), Finland (VTT Technical Research centre of Finland, Keypro), The Netherlands (KWR Water B.V., PWN), Cyprus (Phoebe Research and Innovation, District Local Government Organisation of Limassol), Italy (Engineering Ingegneria Informatica), Germany (Fiware, HST), France (EGM), Belgium (Water Europe), and the United Kingdom (University of Exeter, South West Water).

## Participants

WATERVERSE ecosystem collects various kinds of data concerning water and its management. Participants are thus both public and private entities, e.g. utilities, public administrations, research organizations, large enterprises and SMEs as technology providers contributing to the Water Data Management Ecosystem (WDME), a cutting-edge suite of tools and services aimed at revolutionizing data management in the water sector. Some participants are thus responsible for the delivery of the multi-level architecture of the WDME and its wide spectrum of services, such as data collection, standardization, pre-processing and cataloguing. Security providers have been involved in all the technical processes and in each pilot.

## Data sharing model – architecture

WATERVERSE architecture is not properly a data space, since each pilot has its own platform for data collection, improvement and elaboration. However, the Water Data Management Ecosystem aligns with the Blueprint for creating data spaces proposed by the Data Space Support Centre, aiming to ensuring compatibility with broader initiatives by inspiring to common principles, such as the FAIR ones. The structure of WDME develops in five horizontal layers and one transversal layer for managing security and identification aspects, providing user authentication and cyber-security solutions. These layers concern:

- Data collection, enabling to group data from various sources.
- Interoperability and data harmonisation, aligning data to models and to format compliance, facilitating their sharing.
- Data management, storing and distributing harmonised data to other components.
- Data processing, cleaning, reconciling, anonymising, clustering data to generate added value for decision-makers.
- Application, delivering the processed data to other platforms, e.g. for data visualization or to create interactive dashboards.

## Types of data

WATERVERSE manages a wide variety of data types related to water management, ranging from environmental to infrastructure data. The core of each pilot is data from the local water companies and utilities involved. This is typically supplemented by meteorological and satellite data, mainly from Copernicus, as well as other types of information related to the specific objective of the test. As the data ecosystem of each pilot is therefore different and serves different purposes, the water utilities do not share their data with each other. However, they all meet the same interoperability criteria to facilitate future implementation in a thematic water data space.

## *Interoperability*

WATERVERSE involves members of the technical community which is driving the development of data spaces to foster synergies between these ecosystems and the solutions developed in the project. FIWARE Building Blocks and comprising tools and methods have been adopted to ensure security according to European shared criteria and effective data management. Compliance with technical specifications is considered as necessary to increase water utilities' data interoperability. For this reason, WATERVERSE has aligned with the latest definition of Datasets and Data Services of the DCAT-AP, a DCAT application profile for sharing information about catalogues containing Datasets and Data Services descriptions in Europe as a first step for interoperability Europe.

## *Trust*

Addressing the issue of trust in WATERVERSE meant trying to change the conservative approach of water utilities to data sharing. For this reason, the project emphasizes data security and privacy, ensuring protection and responsible management. Blockchain technology is integrated in the ecosystem to provide transparency and traceability of data transactions. Furthermore, the pilots are aimed at increasing trust in the benefits of data sharing as a way to address local needs and challenges: to this aim, trainings have been developed, helping water utilities to properly use tools and services from the Water Data Management Ecosystem.

## *Present and future sustainability*

WATERVERSE is a 36-month project funded by the European Union's Horizon Europe Research and Innovation programme. Since the end is scheduled in September 2025, the consortium is actually working on an exploitation plan through which ensure continuity to the developed pilots and engage new test areas with a proper business model, based on the adoption of the WDME. By improving the efficiency of water utilities and promoting sustainable water management practices, the project contributes to the long-term resilience of the water sector. The long-term goal of the project partners is to use the experience gained in the pilots and possible other use cases that might be developed to create a European Water Data Space from the infrastructure deployed during the project.

## *Obstacles*

One of the main objectives of the project was to support a change of mindset among water utilities on the issue of data sharing and valorisation. The mistrust of these actors, the fear of losing control of their data and the lack of awareness of the importance of quality data were the main obstacles encountered by WATERVERSE. As a result, the activities of collecting and pre-processing data from water utilities were very time consuming and costly. To speed up this process, tools were developed to reconcile time series of data and identify any extreme values. At the same time, training water utilities in the use of these tools and the benefits they can bring to their services and to better management of water resources in general was crucial. Thanks to a greater awareness of the value of their data, the utilities involved are now keen to continue using the components developed in the project even after its completion.

## Policies

The WATERVERSE project has shown how important it is to demonstrate to water utilities the long-term sustainability of economic investments made to improve data quality in order to share and create value, in order to effectively change attitudes and mindset on data. This can be done through incentives, staff training and, most importantly, the identification and promotion of experiences in the sector that can become good practices for inspiration and comparison. Furthermore, WATERVERSE is currently working on defining some specific policy recommendations based on the pilot experiences.

# USAGE

## Challenge

The main challenge addressed by USAGE **(Urban Data Space for Green Deal)** is to help cities achieving the overall objectives of the European Green Deal, by fostering the double transition (environmental and digital) and supporting the implementation of the European strategy for data. In fact, at urban level, it is necessary to better understand what data and technologies are available and what still needs to be developed for smarter and more sustainable cities. Moreover, scalability and replicability of best practises for multi-player data-sharing are to be fostered in a stronger way, in order to design more effective policy practices and instruments to support evidence-based decision making, thus benefitting especially small and medium-sized European cities, where climate change is mostly felt.

## Solution

USAGE aims to build a decentralised infrastructure for trustworthy collection, processing and exchange of urban data to support public bodies, private entities and civil society in the design and adoption of Local Green Deals (LGD). The project supports this type of data valorisation by combining tools and infrastructures, deploying software solutions for data sharing, creating data governance frameworks for cities (starting from Ferrara, Graz, Leuven and Zaragoza as four use cases) and improving the availability, quality and interoperability of data for LGD. In the context of the USAGE project, the data space is conceived as a dynamic ecosystem where data from different sources are shared (automatically and securely) and analysed together. This approach overcomes the limitations of traditional open data and fosters the creation of innovative analytical models that can transform seemingly random data into meaningful patterns and strategic actions.

## Development stage and use case

The project is currently in the last months of implementation phase and has a total duration of three years. USAGE aimed at supporting cities in adopting local strategies aligned to the Green Deal fostering a digital and sustainable transformation. The data space is thus the endpoint of the project and a possible beginning for new and wider initiatives. The draft of the rulebook is well defined in terms of roles, rules, data, and technical specification, and it will be signed by each partner in April 2025. Four use cases across Europe have been developed in the project, each of them addressing a specific climate challenge: Ferrara (IT) aimed at expanding green spaces, mitigating heat islands and flash flooding events, with IoT and sensor data; Graz (AU) focused on modelling, simulating and predicting urban climate conditions with AI and real-time environmental data; Leuven (BE) used ML to identify buildings suitable for green roofs in order to become climate neutral by 2050; Zaragoza (SP) tried to connect citizen participation, open data and geospatial data to Local Green Deal priorities.

## Geographical area of reference

USAGE solutions are being validated in four pilot areas in Italy (Municipality of Ferrara), Austria (Municipality of Graz), Belgium (Municipality of Leuven), and Spain (Municipality of Zaragoza), each of them typically involving stakeholders at regional level and using their data (e.g. in the Italian use case, Arpae and Emilia Romagna Region are important sources of data, as well as the city of Ferrara). Cooperation at different territorial levels is considered to be strategic in order to both do impact on a larger scale and investigate reusability of local solutions in other urban areas.

## Partners

Coordinated by Universidad Politécnica de Madrid, the consortium involves local government, industry & SMEs, academia & research and standardization organizations. It has complementary expertise among all project partners, aiming at covering most of the types of stakeholders that should be involved in the design of urban data spaces that address some of the Local Green Deal priorities. Project partners are Ayuntamiento de Zaragoza from Spain, Vermessung AVT-ZT-GmbH from Austria, Open Geospatial Consortium Europe, the Lisbon Council and Katholieke Universiteit Leuven from Belgium, GeoCat B.V. from the Netherlands, and AVT Airborne Sensing Italia srl, Municipality of Ferrara, SIPRO Ferrara, Deda Next, Epsilon Italia and Bruno Kessler Foundation from Italy. In order to develop the local pilots, other regional stakeholders, both public and private, have been involved.

## Participants

USAGE ecosystem collects various kinds of data concerning water and its management. Participants are thus both public and private entities, e.g. public administrations, utilities, enterprises, and research organisations. Since the project aims to show to cities the value in connecting digital data tools to create circular value chains in various sectors, actors involved in each ecosystem support the spread of data-driven policies with data related to energy efficiency, air quality, climate condition, water management, use of soil, waste management.

## Data sharing model – architecture

The project has been implemented with tools and modules inspired by the Gaia-X e IDSA approaches and architectures. The aim of the use cases, which adopt emerging technologies of AI, IoT and blockchain, is to prepare a set of technological solutions that can easily interact with the main structures used in operational data spaces. USAGE architecture includes a catalogue of data and tools, a geographic server, and a dashboard:

- The catalogue publishes and facilitates the discovery of metadata of datasets and software, both re-used or developed within the project. Dataset metadata conforming to the INSPIRE and DCAT-AP standards are natively published in the pilot city portals harvested from the catalogue, while metadata for AI algorithms and tools are created directly in our catalogue using an ISO-19115 compliant template.
- GeoServer open source is used to share USAGE data according to interoperable OGC standards. Machine-to-machine interactions are implemented through combined use of the Catalogue and of the geographic server, as the links to the web services are provided in the metadata published in the Catalogue.
- The Indicators Dashboard enables the visualization of indicators defined to measure the successful set-up and operation of an urban data space prototype in each of the four pilot cities involved in the project. Radar graphs enable monitoring the temporal trend of the cities in achieving the target values fixed for each indicator.

## Types of data

USAGE addresses urban challenges by managing different types of geospatial data, such as building information, remote sensing data and air quality measurements. This includes monitoring flash floods and creating digital models of terrain and buildings. Satellite remote sensing provides valuable indices, while data on trees and public green spaces gives an insight into urban nature. The water network layout is also detailed, including nodes, and district heating and sewerage networks are mapped, distinguishing between mixed, white and black networks. Population data and cadastral information help to understand the demographics of the community. Waste management is addressed with the precise location of municipal waste containers and door-to-door collection areas. Public infrastructure data includes light points, public lighting, electrical panels and photovoltaic signals. Traffic systems are dynamically monitored, recording the total number of transits and pedestrian calls per hour, as well as the average hourly background noise. At-risk areas and green spaces are identified for mowing, and a public tree census tracks urban forestry. Parking capacity, both on-street and in facilities, is detailed, including blue spaces and real-time occupancy statistics. Altimetry data, including isolines and surfaces, provides topographical insight, while information on flood zones, urban zones and drainage basins completes the comprehensive overview.

## *Interoperability*

The project consortium aims to implement replicable initiatives and provide solutions that can be easily adopted by other European cities willing to tackle climate change through data valorisation. For this reason, the technological architectures of USAGE use cases are aligned with the main features of Gaia-X and IDSA approaches. The adoption of standards for data products and services (i.e. WMS, WFS, API, OGC) as well as for metadata (which are compliant with INSPIRE and DCAT-AP) addresses the challenge of semantic interoperability.

## *Trust*

To ensure mutual trust and compliance within the ecosystem, USAGE adheres to all relevant data protection regulations and standards, thus providing the necessary data security. Since involving both private, public and research actors, strong commitment by project partners and stakeholders was fundamental. To this aim, rules and methods for onboarding and participation have been widely discussed and shared among the partners since the early stages of the project. The collegial nature of these decisions strengthened internal trust within the consortium and consequently facilitated the involvement of other local actors as data providers in the implementation of the use cases.

## *Present and future sustainability*

The three-year project USAGE is currently funded by the European Union's Horizon Europe Framework Programme for Research and Innovation (call HORIZON-CL6-2021-GOVERNANCE-01-17). Since it will end in July 2025, the issue of sustainability is being discussed: the “pro-bono” collaboration between the founding partners will allow the four use cases to remain operational for a few months after the project closure, while potential business models will be evaluated to ensure continuity and possible scalability / replicability of the developed initiatives.

## *Obstacles*

The main obstacles encountered during the implementation of the project concern the insufficient information about the approach of data spaces for data valorisation and the lack of running examples in the public sector. The spread of this approach, almost only in the private sector, has led public administrations to fear making mistakes or taking risky initiatives. The public sector's ability to activate the entire territorial ecosystem, from research players to businesses, has averted this risk and made it possible to build local partnerships in each of the pilot projects, capable of supporting a change of mentality in the approach to data and the implementation of effective solutions. An important and challenging process of dialogue and collaboration has been carried out with the main stakeholders in order to develop a 'founding document' of the data space defining the roles of the different actors, the data involved, the policies and technical specifications for data exchange and access.

## *Policies*

According to the experiences of USAGE, in order to effectively support the dissemination and diffusion of data spaces, many specific use cases should be developed: this should stimulate interest in the approach and attract thematic providers and consumers who would exploit the opportunities associated with the data space. Furthermore, it would be very useful to have guidelines at national level interpreting the principles and logic of data spaces in the public sector, especially in the Alpine regions, which are very "rich" in regulations at different institutional levels. Policies should thus promote these initiatives as a form of public-private cooperation outside the normal supply and procurement channels.

# AgriCS

## Challenge

AgriCS **(Agricoltura, Conoscenza, Sviluppo)** aims to foster the transfer of knowledge and research experience to agricultural enterprises in an innovative way, seizing the opportunities offered by digital transformation. The project thus addresses the challenges of digitalisation and data valorisation in the agri-food sector of the Italian region of Friuli Venezia Giulia, and faces obstacles related to mistrust and lack of awareness of the functioning and the opportunities of digital tools. Incorporating new skills and instruments would benefit businesses, by promoting their economic growth in line with sustainability values. Moreover, overcoming these challenges is also necessary to effectively address rural development issues, such as climate change, energy consumption, resource efficiency and environmental protection.

## Solution

AgriCS addresses these challenges by taking advantage of ICT technologies, on the one hand as support for information activities and the dissemination of technical knowledge, and on the other hand through a web platform equipped with decision-support tools for agri-food enterprises. The project therefore develops in a process of digitisation, in which entities are first trained on the issues of data collection and valorisation to optimise their business, and then accompanied to use mathematical models for simulation and forecasting made available on the platform by the Friuli Venezia Giulia Regional Agency for Rural Development (ERSA). Covering a wide variety of topics and areas, ranging from phytosanitary defence, irrigation and water optimisation to fertilisation and the simulation of the economic and environmental effects of technical and management choices, these models concretely and effectively support the dual transition, digital and sustainable.

## Development stage and use case

After being supported by the European Agricultural Fund for Rural Development (EAFRD) until 2024, the project continues to operate through regional funds, which ensure the evolutionary maintenance of the platform. AgriCS was first developed in 2018, thus it cannot be inspired by the data space framework that emerged a few years later. Nevertheless, it incorporates some of the basic principles that guide data spaces, by involving different types of stakeholders, focusing on a systemic and data-driven approach to complex challenges, and promoting data valorisation to support economic growth and sustainability. The project's use cases are more accurately described as "models" that are made available free of charge to agricultural enterprises for optimisation and planning purposes: eight models are dedicated to phytosanitary aspects, three to agronomic aspects and six concern agrometeorological scenarios.

## Geographical area of reference

As a regional initiative, the AgriCS project focuses specifically on Friuli Venezia Giulia (FVG), in the north-eastern part of Italy. This region is geographically characterised by the Carnic and Julian Alps in the north, with rocky soils, and more fertile areas in the central hilly plateau and in the coastal plains along the Adriatic Sea, thus making agriculture a strategic element for promoting economic growth, restoring ecosystems, and improving sustainability. The FVG does not typically suffer from water scarcity, but is prone to atmospheric instability, with frequent thunderstorms affecting the characteristics of the soil.

## Partners

The project is led by the Regional Agency for Rural Development (ERSA) and is supported by the Friuli Venezia Giulia region through the Central Directorate for Agri-food, Forestry and Fishery Resources. Key members are the regional in-house ICT company, which ensures the correct functioning of the platform, and the Regional Environmental Protection Agency (ARPA FVG), which collects and provides meteorological data that feeds the AgriCS models.

## Participants

AgriCS aims to promote the digitalisation of agricultural enterprises by supporting the transfer of knowledge from experiments conducted by research organisations, which are thus involved in the project. The AgriCS ecosystem is mainly populated by private entities in the agri-food sector, but participants also include regional public actors providing geographic and meteorological open data.

## Data sharing model – architecture

The AgriCS architecture consists of a platform designed to facilitate a secure interaction between ERSA and agricultural companies, with a public section and a restricted section for business simulations. At the heart of the project are the mathematical models that will be made available to farms and technicians as decision support systems. As it was decided not to outsource the supply of the models, the technical management of the platform is reserved for ERSA technicians, who have access to a test area where they can experimentally test any changes to the parameters of the equations and simulations. The models are based on a soil information layer accessible via a dedicated web-GIS and cover phytosanitary, agronomic and agrometeorological scenarios:

- The eight phytosanitary models concern the cultivation of vines, apple trees and maize and, thanks to the integration of ARPA FVG data, allow the selection of a weather station or the loading of data from the farm's weather station and the generation of forecasts up to the next 3 days, returned with a graphical output and downloadable in JSON format.
- The three agronomic models simulate the effects of the technical choices made by the farm in terms of irrigation and fertilisation of the crops, and the sustainability of the unit in terms of environmental and economic criteria. The platform's web-GIS substrate allows the user to select the parcel of interest to which the soil characteristics are already associated (which can be modified if the user has more recent data), and the simulation provides graphical and tabular output that can be downloaded in JSON format for further processing. The fertilisation model estimates the development of each crop on the basis of organic and mineral fertilisation, taking into account water balance values and possible thermal stresses; the irrigation model provides operational indications and alerts for irrigation interventions, making it possible to assess the impact of agronomic choices on the water balance and to limit crop dry matter losses; finally, the farm management and sustainability model makes it possible to calculate the carbon footprint, the water footprint and to determine the energy balance of the farm (positive in the case of soil organic matter accumulation, negative in the case of its consumption).
- The six agrometeorological scenarios have a macro-territorial character and take into account both phytosanitary and agronomic aspects. Water resources are particularly important in the calculation of the territorial water balance (influenced by variables such as cumulative rainfall and evapotranspiration) and in that of the soil viability index.

## Types of data

The AgriCS models are supported by various types of information needed to develop simulations of agricultural scenarios. Soil data - both public and private, from agri-food companies - feed the web-

GIS platform, with 222 cartographic units and 165 soil types identified. Moreover, meteorological data from ARPA FVG are utilised to forecast relevant phenomena. For this reason, daily or even hourly information from meteorological stations is integrated with radar observations, with a granularity of 500m x 500m. The project also manages data from ERSA and other research centres on irrigation, fertilisation, soil and plant health, which enable the models to properly function.

## *Interoperability*

To integrate data from multiple sources, the project has developed open-code simulation models in the SCALA language, which is interoperable with Java. The AgriCS solution is in line with the national standards for data pooling and sharing that govern the agri-food sector, allowing any company to upload its data to the platform and improve its activities, thus promoting economic growth and environmental sustainability.

## *Trust*

In the AgriCS ecosystem, the issue of trust mainly concerns private entities that feed the models with their data to gain insight into the possible effects of different scenarios. To promote security, authorisation is required to access to restricted areas of the platform. Users thereby need to fill in a form with identity data and information about their relationship with the Regional Rural Development Programme as owners or legal representatives of an agricultural enterprise. This procedure prevents external entities from accessing private data. In addition, AgriCS does not collect the companies' simulation data, nor does it share them with third parties, public or private. Trust among the project stakeholders and beneficiaries is also supported by the direct involvement of ARPA FVG as a key interlocutor for agricultural entities in the partnership.

## *Present and future sustainability*

AgriCS is a regional project activated in 2018 and funded until 2024 by sub-measure 1.2 of the RDP 2014-2020 of the Autonomous Region of Friuli Venezia Giulia with the support of the European Agricultural Fund for Rural Development (EAFRD). After an initial three-year period, the focus of the project has shifted to innovations in the demonstration and digital aspects of AgriCS, with the aim of ensuring its sustainability through dissemination and demonstration activities aimed at raising awareness among agri-food enterprises of opportunities associated with integrating information generated by the models made available on the platform into the decision-making process, thus expanding the community of users. To date, the project is operational and the costs for the evolutionary maintenance of the platform are covered by regional resources. The issue of sustainability has been central from the earliest phases of the project: ERSA's decision not to entrust an external economic operator with the supply of the mathematical models, which would have created an almost total dependence on another actor, was a rewarding one. On the contrary, by owning the models and providing itself with resources with IT skills ERSA was able and will be able in the future to develop new code equations and possibly new declinations of the models independently. The future of AgriCS foresees integration with other initiatives implemented within the Central Directorate for Agri-Food, Forestry and Fishery Resources, starting with the SiCaNSE project (information System for Natural Capital and Ecosystem Services in the Agro-Forestry Sector). The aim of this project, in collaboration with the Universities of Udine and Trieste, is to develop a set of tools both for the territorial analysis of issues similar to those of AgriCS (such as crop systems, soil erosion, soil fertility, irrigation needs, monitoring of agro-ecosystems) and for the valorisation of ecosystem services, starting from the information available in the farm files of agricultural enterprises, which may also be of interest for AgriCS models. The data space approach could be successful in supporting a smooth integration between solutions and facilitating the user experience.

In this way, it would become easier to host in the same space also information currently hosted on the regional platform UBWeb (Forest Web Utilisation) related to forest areas, their monitoring, fire and landslide risk assessment and forest utilisation in terms of quantity of wood cut and used, and to foster the development of data-driven policies guided by a cross-sectoral awareness.

## *Obstacles*

AgriCS addresses the challenge of promoting digitalisation in the agricultural sector, so low digital literacy and lack of awareness of the importance of data for action planning initially hindered the use of the platform and of its models. The project faced the risk of not reaching and involving companies, especially the SMEs, which would have meant failing in its main objective of supporting the transfer of knowledge from research experiments. ERSA therefore developed specific training for agri-food enterprises to learn how to collect data and how to use them in the models, demonstrating the direct benefits of digitalisation. This initiative ensured wider participation in the ecosystem and, through a cascade mechanism, convinced some producer associations to join the platform.

## *Policies*

The project successfully demonstrates the value of data valorisation applied to simulation models in the agri-food sector and represents an example of fruitful private-public partnership. Digitalisation, however, still needs to be promoted, not only through economic incentives, but also with dedicated trainings. The interest of other regions and countries in replicating this experience also leads to consider adopting the data-space approach as a strategy for smooth interaction between different local ecosystems, as is already happening in the Friuli Venezia Giulia region with the fertile dialogue between AgriCS and SiCaNSe. A possible synergy at national or international level would help to define common and data-driven rural development policies.

# datahub.tirol

## Challenge

datahub.tirol tackles the challenges of limited structured access, low interoperability, and poor visibility of regional data. The initiative aims to unlock the potential of data to support digital innovation across the Tirol region, with a focus on key sectors such as tourism, mobility, energy, agriculture, and public administration. To this end, the project aims to serve as a cross-sectoral platform, offering a shared digital infrastructure and acting as a central hub connecting a diverse ecosystem of partners.

## Solution

The datahub.tirol offers a comprehensive solution by combining both technical and organizational infrastructure to enable the structured provision, secure sharing, and effective reuse of data. Its data space relies on open standards to ensure compatibility and promotes a multi-layered governance model to manage data flows transparently. Additionally, it supports data providers through practical guidance and tailored assistance, helping them navigate the evolving landscape of data ecosystems and AI-driven innovation.

## Development stage and use case

datahub.tirol has been operational since late 2023. It has already supported concrete applications, such as the provision of event-related data during initiatives like the digital.tirol Innovation Week. New use cases are currently being developed, expanding the platform's scope and demonstrating its potential across different sectors.

## Geographical area of reference

The datahub.tirol primarily focuses on the Tirol region in Austria, thus addressing core economic activities of this area, and especially tourism. Starting from local data management needs, the project aims to expand across borders within the Alpine region, in line with the goals of the EUSALP (EU Strategy for the Alpine Region) initiative.

## Partners

The initiative is supported by a strong network of actors, including the Regional Government of Tirol, Lebensraum Tirol Holding, Standortagentur Tirol, Tirol Werbung, and Fraunhofer Austria. It is also closely connected to projects such as DIH-West and digital.tirol, and involves a wide range of regional, cross-border, and national stakeholders from the public sector, research institutions, and industry.

## Participants

Datahub.tirol brings together a diverse mix of public and private organizations, including government authorities, tourism associations, research centres, start-ups, and established companies. To date, over 30 organizations are actively involved, each contributing to different capacities across the data ecosystem.

## Data sharing model- architecture

datahub.tirol adopts a pragmatic, user-oriented approach to data sharing, while aligning with the core principles of IDSA (International Data Spaces Association) and Gaia-X. Its architecture is based on modular, open-source components, with a strong emphasis on interoperability and federated data provision. For the connectors to its data space ecosystem, datahub.tirol uses also third-party software solutions from Nexyo.

## Types of data

Since various sectors are addressed by this initiative, lots of different types of data are managed within datahub.tirol, including both public (e.g., Open Government Data) and non-public data (proprietary data from companies or associations). Data availability depends on licensing models and individual agreements.

## Interoperability

To ensure seamless data exchange, datahub.tirol relies on open standards such as DCAT-AP and schema.org. It supports integration with cloud service providers and the use of REST APIs, enabling flexible connectivity. Interoperability is further enhanced through the adoption of metadata standards and the use of data pipelining tools, facilitating structured and scalable data integration across systems.

## Trust

datahub.tirol fosters trust through a combination of multi-level access control, transparent governance, adoption of international standards, and specific trust-building initiatives for entities interested in entering the data space or already involved in its ecosystem. These include workshops, individual consultations, and practical guidelines designed to support participants. Open and consistent communication plays a central role in creating a reliable and collaborative data-sharing environment.

## Present and future sustainability

Currently, the data space is funded by the Regional Government of Tirol and contributions from project partners. Looking ahead, long-term sustainability will be pursued through a mix of service-based offerings, project-driven development, and potential financial contributions from users. There are also plans to expand the model to other regions, reinforcing its scalability and impact.

## Obstacles

Key obstacles encountered include the absence of structured data, legal uncertainties surrounding data sharing, and the need to build trust with data providers. Additionally, the integration of existing systems proved to be technically complex in certain cases, requiring tailored solutions and close collaboration with stakeholders.

## *Policies*

datahub.tirol addresses the need to foster regional development through data valorisation and reuse and with an ecosystemic perspective embracing different sectors. To support the spread of solutions like this, policies should establish clear legal frameworks for data sharing, provide targeted funding for data infrastructure, foster support for standardization and open-source software, and promote digital skills development within public institutions. Moreover, activating funding programmes would be highly advantageous.

# TRUZZT Alliance

## Challenge

The Truzzt Alliance tackles critical challenges concerning data sharing and valorisation, including data sovereignty, security, interoperability, trust anchors, innovation, user experience, and the advancement of an open European data ecosystem. The aim of the Alliance is to create secure, scalable, and standardized digital infrastructures through decentralized data rooms and identity management solutions. Therefore, the Truzzt Alliance addresses critical issues for data spaces implementation from technological providers' perspective.

## Solution

To overcome these challenges, while working on the definition of open standards, the Alliance has created and is now managing data rooms, consistent and secure spaces designed to host data and foster their shareability in a trustful way. This solution thus focuses on:

- Ensuring Data Sovereignty: the Truzzt Alliance utilizes decentralized data storage and exchange mechanisms, ensuring that data remains under the control of its rightful owners. This approach fully adheres to European regulations such as GDPR, IDSA, and Gaia-X, ensuring legal compliance and secure data management.
- Building Trust Anchors: trust is established through cutting-edge technologies like online ID verification, digital certificate certification authorities (CA), and dynamic attribute provisioning services (DAPS). These tools verify identities and enable secure and trusted data exchanges.
- Promoting Interoperability: by following open standards such as IDSA and Gaia-X, the Truzzt Alliance facilitates the seamless integration and scaling of data spaces across different industries, networks, and countries. This promotes compatibility with other European data ecosystems, such as the Mobility Data Space.
- Driving Innovation: by enabling secure and controlled data sharing, the Truzzt Alliance empowers the development of new business models and services. For instance, the Mobility Data Space allows mobility providers to monetize their data, while fostering innovative solutions for future mobility challenges.
- Enhancing Security: advanced encryption methods, transaction monitoring logging systems, and secure cloud infrastructures guarantee robust protection against unauthorized access or misuse of data, ensuring the highest security standards.
- Supporting an Open European Ecosystem: the Truzzt Alliance actively advocates for the widespread adoption of universal standards, strengthening Europe's position in the realm of digital sovereignty. By fostering cross-industry and cross-country collaboration, it plays a vital role in building a sustainable and interoperable European data ecosystem.

## Development stage and use case

The Truzzt's initiative is already in an operational phase. As mentioned earlier, Truzzt has contributed to the development of several use cases, including:

- Mobility Data Space (MDS- https://mobility-dataspace.eu/): a European platform for the secure sharing of data in the mobility sector. It enables companies and institutions to exchange data in a sovereign manner to develop innovative solutions, such as smart transport systems. It is part of the European initiative Gaia-X and will be integrated into the German public platform Mobilithek.
- Datenraum Kultur (https://datenraumkultur.de/): a digital space for sharing and using data in the cultural sector. It involves museums, theaters, archives, and cultural platforms in digitizing

content, improving access to cultural heritage, and creating new digital experiences. It fosters cooperation between public and private cultural stakeholders.

- ZVEH (https://www.zveh.de/): the central association of German craft enterprises in the electrical and information technology sectors. It represents the interests of the industry, promotes training, technological innovation, and the development of smart solutions for buildings and homes. It serves as a key reference for professionals and companies in the sector.
- BEM (https://www.bem-ev.de/): a German association that promotes electric mobility. It supports the transition to a more sustainable transport system through advocacy for environmental policies, charging infrastructure, and innovations in electric vehicles. It collaborates with businesses, policymakers, and research institutions.

## *Geographical area of reference*

The Alliance is primarily involved in European initiatives, but is actually interested in expanding towards Asia, especially in China and Japan.

## *Partners*

The initiative consortium involves a strong network of private, public and research actors, including Acatech, BEM, EVIDEN/Atos, evAI, Ferdinand-Steinbeiss-Institut, Fraunhofer FIT, Hochschule Bonn-Rhein-Sieg, IDSA, IONOS, LCM Linz Center of Mechatronics, MaaS, Mobility Data Space, Smart Living Next, Think.iT, Universität Siegen, Verifeye, WEB ID, ZVEH.

## *Participants*

The various use cases in which the Truzzt Alliance is involved see both public and private entities as participants.

## *Data sharing model- architecture*

Truzzt solution aims to be easily interoperable, therefore it is based on the IDSA Reference Architecture Model and follows the principles and framework of Gaia-X. Sovereignty and user-centricity are core values of the Alliance.

## *Types of data*

Due to the wide variety of participants, both public (e.g., Open Government Data) and non-public data (proprietary data from companies or associations) are managed within Truzzt Alliance.

## *Interoperability*

The Truzzt Alliance focuses on both intra- and cross-data space interoperability by adopting open standards like IDSA and Gaia-X. These standards ensure seamless data exchange within individual data rooms (intra-operability) and across different data ecosystems (cross-operability).Cross-Data Space Interoperability, in particular, is facilitated by strict adherence to IDSA and Gaia-X, ensuring compatibility with other European data spaces like the Mobility Data Space.
The alliance also emphasizes standardization by:

- Implementing IDSA and Gaia-X standards for secure data exchange.
- Offering services to ensure compliance with these standards.

- Collaborating internationally to promote universal open standards.

Truzzt uses established standards like IDS, GAIA-X, and W3C Verifiable Credentials, ensuring both semantic and technical interoperability through standardized APIs and protocols. It also supports interoperable credential management using Self-Sovereign Identity (SSI).

## *Trust*

The Truzzt Alliance ensures trust and compliance by adhering to European standards and regulations. Data is exchanged securely through decentralized, verified participant interactions, with identity verification managed by Certification Authorities and Dynamic Attribute Provisioning Services. Transactions are fully traceable, and digital compliance tools prevent antitrust violations. Secure onboarding and the use of open-source infrastructure enhance transparency, fostering a collaborative environment of trust and accountability.

## *Present and future sustainability*

The creation and development this initiative have been achieved through self-financing and the provision of services to customers. Projects like Mobility Data Space and Datenraum Kultur are supported by dedicated funding, ensuring that both the establishment and ongoing development of data spaces are financially viable. Other data spaces' sustainability is guaranteed by the adoption of a specific business model, such as the one for BEM, which is funded by the BEM itself through membership fees.

## *Obstacles*

The main obstacles faced by the Alliance concerned compatibility with other data spaces, certification requirements, software version compatibility, and acceptance of open-source components. However, by using open standards, decentralized identity solutions, and partnerships, a trusted platform for secure data exchange was successfully built. Additionally, while transferring research projects like Dataspace Culture from prototype to live environments posed some challenges, the main issue was the time and financial investment required, rather than technical difficulties.

## *Policies*

According to the Truzzt Alliance experience, to better support data space initiatives, it is essential to secure long-term public funding, establish clear and harmonized regulations, provide government-backed technical support for open standards, and implement programs that promote ecosystem development and encourage SME participation. Furthermore, active involvement from the public sector as role models and ongoing dialogue between all stakeholders are crucial.

# Policy Recommendations

## Fostering the Data Space Approach in EUSALP

Interreg
Co-funded by the European Union
Alpine Space

Support EUSALP

EUSALP
EU STRATEGY FOR THE ALPINE REGION

# RECOMMENDATIONS TO SUPPORT THE DIFFUSION OF DATA SPACES IN EUSALP

In view of the study carried out and the interviews conducted, the following policy recommendations are intended to foster the implementation of data spaces in order to promote the resilience and growth of Alpine regions.

The principles of interoperability, trust and data sovereignty that characterise this approach support the creation of ecosystems capable of generating value also for the numerous small and medium-sized enterprises and local public administrations in the Alpine territory. It is also for these reasons that greater dissemination of these models will be strategic for the capillary implementation of the European Data Strategy and for increasing the competitiveness of the Union and its macro-regions. Finally, as the research and the interviews have shown, data spaces are frequently the solution to a question that has not yet been asked: the possibility of sharing one's data and products in a secure way, within a decentralised environment where a large number of different actors can enter and each participant has to comply with the same rules, can become a crucial element in supporting innovation in all fields and sectors.

In order for Data Spaces to reach their full potential and for EUSALP to promote their implementation, the following **9 recommendations** are proposed. Of these, five are of a more general and cross-cutting nature, and four are of a sectoral nature – specifically, two for the Manufacturing sector and two for the Water Management sector.

**R1. Support digitalisation as an enabling factor and access condition to the data space ecosystem.**

**R2. Promote the standardisation of experiences and their alignment with European policy guidance through the adoption of existing strategies and tools, and the valorisation of key learnings from already implemented projects.**

**R3. Foster the dissemination of knowledge and the growth of skills.**

**R4. Establish services to facilitate access to the data space ecosystem.**

**R5. Evaluate targeted investments in support of data spaces.**

**R6. Ensure a high level of data security and protection.**

**R7. Encourage the increase of use cases involving small and medium-sized enterprises in the Alpine region in the manufacturing sector through a dedicated technological infrastructure.**

**R8. Improve the quality, quantity, and accessibility of water-related data.**

**R9. Promote a unified approach to water management at Alpine level through a dedicated technological infrastructure.**

## Cross-sectoral Policies

### *R1. Support digitalisation as an enabling factor and access condition to the data space ecosystem*

The territorial, economic, and political context of reference for EUSALP is characterised by significant fragmentation: small and medium-sized enterprises and small local public administrations prevail. The digital maturity level of these entities is often inadequate to face challenges related to data valorisation and access to complex ecosystems such as data spaces.

It is therefore necessary, in continuity with European programs (e.g. Digital Europe) and initiatives (e.g. European Digital Innovation Hubs – EDIH), to continue promoting the digitalisation of businesses and local authorities, upon which the economic competitiveness and attractiveness of the Alpine region depend.  The allocation of funds and financial resources to digitalisation is thus a prerequisite for any further reasoning on the topic of data use and is essential for pursuing the objectives of the European Data Strategy, in which data spaces are considered a tool to facilitate the collection and sharing of data for the benefit of citizens, business, research institutions, and public administrations throughout Europe.

Supporting SMEs in their digitalisation process will ensure more equitable access to data spaces, which today are often poorly participated in by these types of enterprises and are instead led by large companies, which are more sensitive and expert in leveraging their data. The data space approach, inspired by the principle of decentralization, represents a valuable opportunity for economic growth for Alpine SMEs that have achieved a good level of digital maturity.

### *R2. Promote the standardisation of experiences and their alignment with European policy guidance through the adoption of existing strategies and tools, and the valorisation of key learnings from already implemented projects*

The data space initiatives to be simulated within the Alpin region context must contribute to the objectives of the European Data Strategy, and in particular to the creation of a single data market guided by the principles expressed in standards and regulations in force at European level. EUSALP policymakers are therefore called upon to work together to ensure full compliance with these cross-cutting regulatory criteria, in addition to those specific to each sector and domain in which data space experiences and use cases may be developed.

In complement to this form of alignment, it is necessary to support the standardization of future initiatives for the interoperability that guides the data space approach and allows an organisation not to limit its participation to a single use case, but to become part of multiple ecosystems. This objective requires pursuing a twofold strategy: on the one hand, it is necessary to promote, at the political level, the definition of a single and as open as possible standard for the architectures and technological layers on which data space are implemented, also in order to avoid - as  the DSSC also argues - " reinventing the wheel" in a context that is already technically and conceptually very complex; on the other hand, it is essential to bring to value the processes implemented, the good practices to be scaled and replicated, the evaluation metrics measured, and the lessons learnt from data space experiences underway or already completed within the broader European context.

The mapping of such case studies is not easy, as this paper shows, due to high fragmentation and the lack of a unanimous agreement on the definition of what data spaces initiatives are. The use of tools such as the Data Space Radar must therefore be complemented by strong synergy between regional policymakers, with the aim of increasing outreach in the search for existing experiences and defining a catalogue that is as exhaustive as possible, not limited to the EUSALP territory but also able to bring other European cases to value. This effort will make it possible to identify guidelines and successful models that can guide future economic investments aimed at supporting the development of new experiences.

## R3. Foster the dissemination of knowledge and the growth of skills

The topic of data spaces is of recent origin and available literature on it is very limited, to the extent that – as previously mentioned – even agreement on terminology is not yet sufficiently consolidated. Support for the digitalisation of small private and public entities and the mapping and cataloguing of completed experiences must therefore be accompanied by an effective action of communication and promotion of the possible strategies for the valorisation of data and, in particular, of data spaces, in order to qualify the EUSALP territory as the leading macro-region at European level on these issues and processes.

The dissemination of a solid data culture also involves training and upskilling activities: such educational pathways are fundamental both for the staff of low-tech companies and for that of public administrations, which are often undersized and suffer from low generational turnover. Increasing digital literacy across all sectors is crucial in today's scenario of rapid technological innovation and can be a determining factor for the creation of new data spaces and related use cases and, on a macro level, for the economic development of the Alpine regions. The provision of qualified training will benefit from tools and contents already developed in the European context, such as the Academy on data.europe.eu and the Blueprint of the Data Space Support Centre, which has been referred to for the theoretical description of data spaces, as well as from funding lines within major programmes such as Interreg, Horizon and Erasmus+.

To support the decision-making process that could lead public and private entities to consider the creation of or participation in a data space, a concise **checklist** has been developed, which is attached to this document, with a series of questions designed to better assess the suitability of a data space and to be prepared for its potential implementation.

## R4. Establish services to facilitate access to the data space ecosystem

Digitalisation is a prerequisite for public and private actors to be able to evaluate any strategy for data valorisation. However, even when a greater data culture has spread within an organisation, actual participation in a data space may still require additional forms of support. For this reason, it may be appropriate to provide specific tools that facilitate access to these ecosystems, also for smaller entities.

The main areas on which such services should focus concern:

- Compliance with the technical standards required by the ecosystem to be accessed: entities could benefit from alignment with the interoperability standards required by the data space and thus be able to create value from "clean" and accessible datasets and data products.
- Verification of compliance with the main regulations in force at the sectoral, local, national, and European level: in this way, entities will be supported in the analysis of the standards to be complied with and the overall trust in the data space instrument will be strengthened.
- Measuring impact: in addition to technical-technological support, it is necessary to indicate to entities wishing to enter a data space which metrics will be taken into account for evaluating the impact of the use cases implemented. These indicators, which may of course vary

depending on the sector, will be identified from the analysis of existing good practices in the wider European context, and will always need to be aligned with the criteria and advancements required to ensure that experiences in the EUSALP area are consistent with the European framework for impact evaluation.

- Risk assessment: the issue of trust does not only concern ensuring that all participants in a data space comply with the same technical and regulatory access criteria but also requires supporting interested entities in evaluating the risks associated with any investment. This element is particularly important for data spaces, as this approach is not yet widespread, is rather complex and the benefits may not be immediately apparent. Such support may therefore be very useful to increase trust in these ecosystems and to populate them with a wider range of actors, thereby enhancing the potential for economic growth and the development of new services.

## *R5. Evaluate targeted investments in support of data spaces*

The promotion of data spaces as part of the European Data Strategy was immediately accompanied by R&D activities carried out by research institutions, large public organisations, and major private players in the technology sector, with the aim of building technological architectures, business models, and governance tools suitable for these ecosystems. To date, the main actors engaged in defining replicable standards and processes, as already highlighted, are in ongoing dialogue to ensure consistency of experiences, interoperability of architectures, and accessibility to the tools necessary for the effective implementation of data spaces.

For this reason, and in consideration of the characteristics of the Alpine territory, where SMEs and small public administrations predominate, it is proposed to the EUSALP policymakers to follow the development of interactions at the level of scientific research, in favour of investments oriented towards the four previous general recommendations, as they are intended for preliminary and indispensable activities for any data valorisation strategy. Once the competitiveness of the Alpine arc ecosystem has been strengthened in this way, it will then be possible, on the basis of the standards identified at European level, to adopt specific measures to support Manufacturing, by allocating resources to help small and medium-sized enterprises implement new use cases in existing data spaces, and Water Management, by realising a flagship project on water resources, as proposed in the thematic policies below.

# Policies for the Manufacturing Sector

## R6. Ensure a high level of data security and protection

In the Alpine region, the manufacturing sector is mainly composed of small and medium-sized enterprises, distributed along the entire value chain of a product. These companies may therefore hold many types of sensitive data, ranging from employee-related information to data on consumers and end buyers, technical production data, patents and product specifications, and financial information. Compliance with European data protection regulations is now a widespread requirement and a barrier to access to data spaces. However, in order to increase trust in these ecosystems and in the actors – mainly industrial, in the case of thematic experiences on manufacturing – that populate them, it may be useful to envisage strengthening security measures as an additional incentive to participate for smaller businesses that may lack the internal technological resources to monitor these issues on a regular basis.

For initiatives to be developed in the EUSALP area, it may therefore be considered, in dialogue with the governance authorities, to promote the adoption of internationally recognized security certifications, such as ISO/IEC 27001, and to request that among the services available within the data space, continuous monitoring systems for threats to the ecosystem and periodic security audit programmes to identify and promptly address any vulnerabilities are always made available.

## R7. Encourage the increase of use cases involving small and medium-sized enterprises in the Alpine region in the manufacturing sector through a dedicated technological infrastructure

The analysis conducted in the study revealed the presence of significant experiences, within the broader European context, of data spaces and related initiatives in this sector, which is one of the 12 strategic areas targeted by the Common European Data Spaces. These case studies have often been implemented under the leadership of large global industrial players who, by holding the governance of the ecosystems, have often established entry conditions that are not easily accessible to smaller entities. This phenomenon could become a major disincentive for the participation of such companies, which, on the other hand, are not able to economically sustain the investment (and related risk) that would be required to build a separate data space.

To address the need to increase the number of successful use cases, which, according to the results of the interviews conducted, is most important element for the widespread adoption of this approach, it is suggested that the entire EUSALP ecosystem support the creation of the technological infrastructure on which to build a sectoral data space, dedicated to the world of manufacturing and industry, with a focus on small and medium-sized enterprises. The objective of this initiative is to significantly reduce the economic burden that these companies would otherwise have to bear, thus encouraging their participation and the realisation of use cases even with a very short production chain.

The coordinated action of the EUSALP Action Groups will ensure neutral governance of the ecosystem, fully in line with the principles and technological standards adopted in the European context and will increase trust among businesses. This way of promoting and disseminating the approach will be an important incentive for actors at risk of exclusion and will guarantee the creation of new value: members of the data space will be required to adhere to the same constraints, rules, and standards to access the ecosystem, and will therefore be in the best position to share their data and services not only with the actors involved in their specific use case, but also with other participants, thus creating new business opportunities.

## Policies for the Water Management Sector

### R8. Improve the quality, quantity, and accessibility of water-related data

Water management is one of the cross-cutting priorities of the EUSALP macro-strategy and it needs to be addressed by private actors, public administrations, and research organisations, each with a different sensitivity regarding data valorisation. As the interviews also revealed, one important obstacle to the implementation of data spaces related to this resource is often the limited availability of data, its obsolescence, and its incompleteness. In the case of water, the need to promote the widest possible dissemination of a data culture as an essential tool for informed decision-making and long-term planning for the use of the resource is accompanied by the need for joint action by the entire EUSALP group to improve both the type and the amount of available information.

Regional policymakers must therefore support the widespread adoption of advanced technologies and sensor systems for the real-time monitoring of the resource and establish common transnational standards for data collection and sharing, ensuring interoperability between the various local systems and platforms, in order to benefit from data space initiatives. Moreover, especially in relation to public actors, it is essential to monitor the quality and accessibility of open data on water and to implement verification and validation processes, providing entities with tools that report and possibly correct any anomalies. The combination of these actions, agreements, and incentives will enable the private and research sectors to access, on existing platforms and within future data spaces, data that is better suited to the development of innovative solutions for water management.

### R9. Promote a unified approach to water management at Alpine level through a dedicated technological infrastructure

The cross-sectoral nature of this topic has resulted in a significant fragmentation of experiences and implementations of data spaces related, more or less directly, to this resource. Unlike what happened in the case of manufacturing and 11 other sectors, the European Data Strategy did not foresee a dedicated water-thematic Common European Data Space. Water is a resource that cannot be attributed to a single sector, and although the main experiences – both mature and emerging – of data space initiatives mapped during this study focus mostly on the agri-food domain and urban management, it cannot be overlooked that water is fundamental, for example, in industrial processing, from the textile to the energy and steel sectors, as well as in climate modelling and monitoring of extreme phenomena.

The data space approach is to progressively overcome the sectoral nature of current experiences to reach a single, fully interoperable macrosystem at the European level, through which it is intended to build the single market for data. For this reason, it is suggested that the entire EUSALP ecosystem carefully monitor data valorisation experiences related to this issue in order to promote agreed standards and to evaluate the possibility of implementing a lighthouse project on water management, supporting investment in a technological infrastructure upon which a thematic data space could thus be based. The goal of this innovative initiative is to provide greater unity to the numerous initiatives (not strictly data space) spread across the Alpine arc on the topic of water, offering a regulated digital environment in which they can easily become use cases, limiting access costs and promoting the sharing of data and services with other public and private participants, all bound by the same rules.

The realisation of this project may be considered strategic for pursuing EUSALP's cross-cutting thematic priority and concretely stimulates transnational and cross-sector cooperation as the only successful approach to effectively and efficiently manage a resource as crucial as water.

# Annexes

Interreg Alpine Space
Co-funded by the European Union
Support EUSALP
EUSALP EU STRATEGY FOR THE ALPINE REGION

# ANNEX 1 – MAPPING CASE STUDIES ON DATA SPACE INITIATIVES

The following template was created to conduct semi-structured interviews with the identified data space initiatives. The model is developed around thirteen thematic questions that are essential to get a complete overview of the mapped data sharing and valorisation experiences, and it highlights some crucial issues to assess the affinity with the data space approach. The interviews presented in the second part of the study follow this pattern.

**Case Study Name**: ______________________________

| TOPIC | QUESTION |
|---|---|
| **Challenge** | Which challenge(s) are you addressing with your data space? |
| **Solution** | Which solution(s) did you find to the challenge(s)? |
| **Development stage and use case** | Is the data space operational? Insert link to eventual use cases |
| **Geographical area of reference** | Which countries/regions are involved in your data space? |
| **Partners** | What partners were involved in the development of this data space? |
| **Participants** | Which kind of participants are involved in your data space? Are they public or private entities?<br><br>How many? |
| **Data sharing model - Architecture** | Which is the architecture of your data space (e.g. Gaia X, IDSA…)? |
| **Types of data** | Which kind of data is hosted in your data space (private, public, proprietary, non-proprietary…)? |
| **Interoperability** | How do you approach the theme of interoperability (both intra and cross-data space)?<br><br>Are there standardization processes in your data space? |

| | |
|---|---|
| **Trust** | What mechanisms did you implement to ensure trust and compliance?<br><br>Belonging to the data space has fostered trust within the data space ecosystem? |
| **Present and Future Sustainability** | How is the data space funded? (e.g. European project…)<br><br>Have you already discussed how to ensure economic sustainability in the future?<br><br>How do you plan to evolve your data space? |
| **Obstacles** | Which are the main challenges you faced while developing your data space? |
| **Policies** | What would you ask to government officials in order to better support these initiatives (e.g. funding programs, regulations, technical support…)? |

# ANNEX 2 – PRELIMINARY QUESTIONS CHECKLIST FOR DATA SPACE INITIATIVES

The following checklist includes two groups of preliminary questions that entities interested in implementing a data space should answer prior to or during the initial project implementation phases. The first list concern more general questions about the main features of these ecosystems and should help discerning whether a data space is a good solution to the identified challenge. The second list, instead, delves deeper into data spaces, focusing on specific elements described in the building blocks. The primary objective of this checklist is to provide a framework for enterprises and public entities in the Alpine regions to develop new use cases and directly ascertain the benefits of data spaces for economic growth.

## General questions: opting for a data space

- What is the **problem to solve**, and why is the data space the right solution? (If it can be resolved by establishing a single API for example, it is most likely not worth it to set-up a data space, unless there are additional reasons).
- Is there already a **data space available** to join to achieve these goals?
- Making a quick calculation, is there enough **value**/a high enough margin for the infrastructure of a data space to be financed?
- Is the issue that the data space is addressing important enough for the **stakeholders** which will be involved? (as a rule of thumb, if it is outside of the top 3 problems for a stakeholder, it might be difficult to convince them to join the effort)

## Thematic questions: implementing a data space

| TOPIC | QUESTION |
|---|---|
| **Principles, Scope, Objectives** | What is the thematic scope of the data space?<br><br>*Decline the scope in specific objectives, growth aspirations, and profit goals.* |
| **Use Cases, Data Providers and Data Users** | Which use cases should be prioritised, and which will be developed in the future?<br><br>How does the data space initiative generate revenue from use cases and what benefits do they offer to the actors involved?<br><br>*Identify the purpose or problem (business, societal, and/or environmental value) to be solved and actors to be involved (both as data providers and data recipients) in these use cases.* |

| | |
|---|---|
| **Offerings, Providers and Value Creation** | What types of data products and services does the data space have and wish to offer? Who is providing them? What value-creation service(s) should be implemented? How does the data space initiative build revenue from service providers?<br><br>*Define what rules should govern the offerings (e.g. concerning metadata, licensing conditions, quality, standards, etc.). Determine which criteria service providers (on the behalf of the data space or third parties) must meet and how can the data space attract these entities and incentivize developing and offering high-quality data products and services.* |
| **Network Effects** | How does the data space aim to establish critical mass and network effects in use cases, data providers, data participants, and service providers?<br><br>*Same-side network effects arise when the presence of, for example, use cases attracts more use cases. Cross-side network effects occur when, for example, the availability of a set of value-added services attracts additional use cases or participants.* |
| **Governance Authority Bodies and Members** | Who are the participants in the governance authority? Who is eligible to join? What criteria should they meet?<br><br>*Define how the composition of the governance authority should be prescribed, which activities will the governance authority perform, and what processes will ensure the performance and the monitoring of its members' duties and responsibilities.* |
| **Stakeholders** | Does the data space initiative need support from additional organizations, such as industry associations, governmental agencies, or non-governmental organizations?<br><br>*List the stakeholders affected directly and indirectly by the data space or in a specific use case.* |
| **People, Resources and Activities** | What staff, resources, and activities are deployed by the data space to maintain operations? What operational costs does the data space incur?<br><br>*The revenues and costs must align with the data space profit and growth strategies. The main cost factors include development costs, operational costs for data space enabling services, and costs related to governance of the data space.* |

| | |
|---|---|
| **Dynamic Capabilities** | How does the data space keep track of necessary changes in the business model?<br><br>*Follow developments within the data space environment to determine whether and which changes in the business model are needed.* |
| **Trust Framework** | Which existing regulatory schemes and standards are to be considered as a basis/component for the definition of the data space trust framework? Which requirements should be satisfied for registering data provenance & traceability?<br><br>*Define what kind of trust anchors should the data space have to validate and verify the credentials of participants, what kind of identifiers will be used in the data space, and what other attestations (e.g. compliance with policies or specific regulations) will be required for membership.* |
| **Interoperability** | How will the data space manage semantics for its data products?<br><br>*Decide which data element and concept are required to be semantically expressed, in order to evaluate the suitability of already existing data models for ensuring semantic interoperability in the ecosystem.* |
| **Legislative Framework** | How does the data space manage legal aspects and policies? Which legal agreements are necessary for the proper functioning of a data space?<br><br>*Ensure compliance to legislative frameworks and triggers, such as privacy policy, data protection policy, and cybersecurity policy. Define legal arrangements for the foundation of the data space and its organisational form, data sharing and usage rights.* |