This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101135988

AI-Enabled Data Lifecycles Optimization and Data Spaces Integration for Increased Efficiency and Interoperability

Project acronym:	PLIADES
Project name:	AI-Enabled Data Lifecycles Optimization and Data Spaces Integration for Increased Efficiency and Interoperability
Grant Agreement No:	101135988
Call:	HORIZON-CL4-2023-DATA-01
Topic	HORIZON-CL4-2023-DATA-01-02
Type of action:	HORIZON Research and Innovation Actions
Start of project:	01/01/2024

Deliverable 2.1

SoA on data spaces & secure information sharing

Work Package:	**WP2 – User Requirements and Specifications**
Task:	**T2.1**
Lead Beneficiary:	**EURECAT**
Due Date:	**30/09/2024**
Submission Date:	**10/10/2024**
Deliverable Status:	**Final**
Deliverable Type:	**Report**
Dissemination Level:	**PU**

Authors

Surname	First Name	Beneficiary
Khandpur Singh	Ashneet	EURECAT
Ortiz Sánchez	Marcel	EURECAT
Garnica Caparros	Marc	EURECAT
Castellvi	Silvia	IDSA
Dalmolen	Simon	TNO
Tsiakas	Kosmas	CERTH
Nasoulis	Christos	CERTH
Rodríguez	Silvia	INNOVALIA
Aguayo Velasco	Asier	TECNALIA
Cabello	Daniel	DENN
Perez Lopez	Borja	UC3M

Reviewers

Surname	First Name	Beneficiary
Pérez Rastelli	Joshué	CEIT
Vitani	Anastasia	ATLANTIS Engineering SA

Version History

Version	Date	Modifications made by
0.1	15/07/2024	EURECAT
0.2	1/09/2024	EURECAT
0.7	23/09/2024	EURECAT
1.0	10/10/2024	EURECAT after reviewer's feedback

Disclaimer

Executive Summary

This deliverable provides a comprehensive analysis of various data space architectures, specifically IDS-RAM, GAIA-X, FIWARE, and IHAN, using a structured methodology to evaluate their characteristics, strengths, and limitations. Data spaces are collaborative environments where data is exchanged across organizations with agreed-upon rules, and they are emerging as crucial infrastructures for secure and efficient data sharing in the digital economy. Each data space is reviewed using a methodical process that considers functionality, interoperability, scalability, and security, among other features.

In addition, the deliverable also covers other significant data sharing initiatives that complement the data space landscape. Through comparison, the unique characteristics of each data space are highlighted, while identifying commonalities and potential synergies that can contribute to enhanced interoperability and collaboration across different sectors and industries. The security frameworks of these initiatives are compared to evaluate how they address privacy protection and data safety in increasingly interconnected environments.

Security and privacy are critical concerns in the realm of data spaces. Vulnerabilities that could impede the broader adoption of data sharing initiatives are identified. Recommendations for enhancing security protocols and privacy protection mechanism are proposed, which are essential for building trust in data sharing environments.

Finally, advancements in current frameworks are identified, along with a set of proposed extension requirements designed to enhance the capabilities of data spaces. These suggestions focus on the need for standardized governance, improved interoperability, and mechanisms to ensure data sovereignty and ethical data use. By addressing these areas, the deliverable aims to support the development of more robust data-sharing infrastructures that can adapt to the evolving landscape of data management.

This thorough analysis provides insightful information for those involved in the development and uptake of data spaces. It contributes to the advancement of the digital economy and the effective use of data for innovation and growth by serving as a guide for future initiatives aimed at creating more cohesive, secure and efficient data-sharing environments, ultimately contributing to the development of the digital economy and the efficient use of data for innovation and growth, with a strong focus on maintaining the highest standards of security and privacy protection.

Table of Contents

List of Figures

List of Tables

List of Terms and Definitions

Table 1 Definitions

Abbreviation	Definition
ABAC	Attribute-Based Access Control
AI	Artificial Intelligence
API	Application Programming Interface
BDVA	Big Data Value Association
DCAT	Data Catalog Vocabulary
DGA	Data Governance Act
DID	Decentralized Identifier
DSA	Digital Services Act
DSBA	Data Spaces Business Alliance
DSGA	Data Space Governance Authority
DSC	Data Sharing Coalition
DSSC	Data Spaces Support Centre
EDC	Eclipse Dataspace Components
EU	European Union
GDPR	General Data Protection Regulation
GXDCH	Gaia-X Digital Clearing House
IAM	Identity and Access Management
IDSA	International Data Spaces Association
IDS	International Data Spaces
IDS-RAM	International Data Spaces Reference Architecture Model
IHAN	International Human Account Network
ISO	International Organization for Standardization
IT	Information Technology
NGSI	Next Generation Service Interface
ODRL	Open Digital Rights Language
OAuth2	Open Authorization 2.0
PII	Personal Identifiable Information
PSD	Personal Services Directory
REST	Representational State Transfer
RBAC	Role-Based Access Control
SLA	Service Level Agreement

SAML	Security Assertion Markup Language
TLS/SSL	Transport Layer Security / Secure Sockets Layer
UI	User Interface
W3C	World Wide Web Consortium
XACML	eXtensible Access Control Markup Language
XML	eXtensible Markup Language

Preface

The increasing need for secure and collaborative data exchange across various industries has fueled the emergence of data spaces as a key technological solution. Within the European Union, fostering a robust data economy relies on enabling seamless data sharing while upholding strict security and privacy regulations.

Objectives

This deliverable aims to provide a comprehensive understanding of data spaces architectures. This involves explaining the fundamental concepts and principles that underpin these architectures, including their design principles and individual components. Furthermore, we will conduct a thorough analysis of prominent European data spaces architectures, examining their structural framework, functionality, and real-world implementation.

A core objective is to provide insights into how these architectures can be improved to facilitate secure, private, and interoperable data sharing within the European data space landscape. Additionally, the document strives to identify potential security and privacy gaps inherent in these architectures and propose effective measures to address these vulnerabilities. Furthermore, it aims to evaluate the interoperability of different data space architectures, assessing their capacity to seamlessly exchange data and collaborate across diverse systems.

Finally, it seeks to explore potential advancements and extensions to existing data space frameworks, proposing innovative solutions to enhance their functionality and address any identified shortcomings.

Scope

In terms of scope, this deliverable will focus on data space architectures designed for secure and collaborative data exchange across various industries, examining concepts such as decentralization, interoperability, and data sovereignty. We will employ a structured methodology to analyse these architectures, considering factors such as scalability, flexibility, and regulatory compliance.

Moreover, the deliverable will address the security and privacy implications of these architectures, identifying potential risks and proposing mitigation strategies. We will explore challenges related to interoperability, including technical, semantic, and governance issues.

Finally, considering emerging technologies and evolving regulations, the report will recommend potential advancements and extensions to existing frameworks. These recommendations will include a set of extension requirements designed to enhance interoperability, security, and privacy, while ensuring compatibility and resilience in the ever-evolving digital landscape.

1 Introduction

In today's interconnected world, the ability to securely share and manage data across diverse sectors is becoming increasingly crucial. As industries rely more heavily on data-driven processes, the demand for robust, secure, and interoperable data exchange solutions has grown exponentially. This shift underscores the need for well-architected data spaces, particularly within the European Union, where the digital economy is a critical driver of innovation and growth.

The European data strategy emphasizes the importance of creating a cohesive data ecosystem, where data can be shared seamlessly while adhering to stringent security, privacy, and sovereignty requirements. Data spaces represent a fundamental component of this ecosystem, serving as structured environments that facilitate the secure and collaborative exchange of data between various stakeholders. These spaces enable organizations to share data while maintaining control over its usage, thereby fostering trust and collaboration across borders and industries.

The architecture of data spaces is pivotal to their effectiveness, as it dictates how data is stored, accessed, and exchanged, while also ensuring compliance with European regulations such as the General Data Protection Regulation (GDPR).

This deliverable is dedicated to providing a comprehensive exploration of data space architectures, with a particular focus on those implemented within the European context. Our aim is to elucidate the core principles that underpin these architectures, including their design philosophies, structural components, and operational mechanisms. By offering a detailed analysis of prominent European data space architectures, this document aims to shed light on how these systems are designed to support secure and private data sharing, while also facilitating interoperability across different platforms and industries.

Moreover, this deliverable seeks to critically examine the existing data space frameworks to identify potential areas for enhancement. Security and privacy are paramount concerns in any data exchange environment, and this document will explore the inherent vulnerabilities within current architectures, offering practical recommendations for mitigating these risks. Interoperability is another critical factor, and this report will assess the extent to which different data space architectures can effectively communicate and collaborate, identifying barriers to seamless data exchange and proposing solutions to overcome them.

In addition to evaluating the current state of data space architectures, this deliverable also looks towards the future. As technology and regulatory landscapes evolve, so too must the frameworks that support data exchange. This report proposes potential advancements and extensions to existing data space architectures, with a focus on enhancing their scalability, flexibility, and compliance with emerging regulations. By providing a forward-looking perspective, this document aims to contribute to the ongoing development of data spaces that not only meet current needs but also are resilient enough to adapt to future challenges.

Ultimately, this deliverable is intended to serve as a comprehensive resource for stakeholders involved in the design, implementation, and governance of data spaces. By offering insights into both the strengths and weaknesses of existing architectures, as well as proposing concrete recommendations for improvement, this document aims to support the ongoing effort to build a robust, secure, and interoperable data ecosystem within the European Union and beyond.

2 Methodology

In this section, a structured methodology designed to assess key components and functionalities to comprehensively evaluate data space architectures is presented. The methodology outlined below serves as a systematic framework for analysing and comparing different architectures based on predefined criteria.

Conceptual foundation

Our approach begins with understanding the conceptual foundation of each data space architecture. This entails identifying the general objectives, target use cases, and data model focus that underpin the architecture's design. By outlining the fundamental principles guiding its development, a clear framework for further analysis is established.

Governance mechanisms

Next, the governance mechanisms that are integrated into each architecture are explored. Through the evaluation of governance mechanisms, the capacity of the architecture to guarantee data security, integrity, and regulatory compliance is determined.

Security features

An essential component of the analysis are security features, which include measures implemented to protect data confidentiality, integrity, and availability. In this step, vulnerability management techniques, access controls, authentication mechanisms, and encryption protocols are evaluated. By thoroughly reviewing security measures, the architectures resilience to potential threats and vulnerabilities is assessed.

Standardization

The process of evaluating standardization involves determining how closely each architecture adheres to established frameworks, protocols, and data formats. By ensuring alignment with standardized practices, the architecture's ability to facilitate seamless integration and interoperability is evaluated, with particular focus on how well it supports interoperability protocols and complies to industry standards.

Added services

The analysis will explore the availability of additional functionalities offered by the architecture. This includes investigating if the architecture provides tools for data anonymization, functionalities for data quality management, and auditing capabilities for tracking data access and lineage.

Auditing and lineage tracking

Auditing capabilities are essential for ensuring accountability and compliance with regulations. By tracking data access and lineage, the architecture can demonstrate a clear audit trail of how data has been used, modified, and shared within the data space.

Data sovereignty

The architecture's approach to ensuring that data providers maintain control over their data is investigated. This includes understanding the mechanisms employed to enforce data usage policies and manage consent from data subjects. By addressing data sovereignty concerns, compliance with relevant legal and regulatory frameworks are addressed and maintained.

Interoperability

The methodology evaluates the architecture's capability for inter and intra data space interoperability, assessing its ability to facilitate seamless data exchange and integration across disparate environments.

Scalability

The focus will be on understanding whether the architecture can handle increasing data volumes and user demands while maintaining efficient performance. How the architecture adapts to evolving data needs and scales to accommodate future growth will be explored.

Regulatory compliance

The analysis will assess the architecture's adherence to relevant data privacy regulations such as GDPR. We will examine how the architecture supports data governance practices that comply with legal requirements.

Communication protocols

The methodology investigates the protocols [1] used for communication within each architecture, assessing their suitability for secure and efficient data exchange. This involves analysing communication protocols, messaging standards, and data transmission mechanisms.

Development stage

Throughout the analysis, the current stage of development of each architecture is considered, providing insights into its maturity and readiness for deployment. This involves assessing the completeness of features, scalability of functionalities, and level of community adoption or industry support. By contextualizing the architecture's stage of development, decisions regarding its implementation and future evolution are informed.

3 Data spaces architectures

Data spaces are data infrastructures and governance frameworks designed to facilitate widespread data pooling and sharing across organizations and sectors [2], [3]. These architectures provide the necessary tools and services for data processing, data sharing, and federating energy-efficient cloud-capacities [4]. They also establish transparent and fair data governance structures in compliance with EU legislation to enhance the availability, quality, and interoperability of data across various sectors. This section will provide an overview of the design principles that underpin data spaces and a closer look at the European data spaces initiative.

3.1 Design principles

The design principles for European data spaces, as outlined in the document [5], are as follows:

1. **Data Sovereignty**

   Data sovereignty is the capability of a natural person or a corporate entity for exclusive self-determination over their data, i.e., the ability to decide at all times which data is used, who uses it and for what purpose [6].

   Data sovereignty is achieved by ensuring that data owners have full control over their data, including how it is accessed, shared, and utilized. This principle is reinforced by procedures that allow the owner to retain ownership of the data while allowing others to access it in a secure and regulated manner.

2. **Data level playing field**

   This principle ensures that participants in the data space compete based on the quality of their services rather than the quantity of data they control. It is fundamental for creating a fair data exchange economy.

   Achieving a level playing field entails developing open standards and protocols that provide equal access to data resources and transparency in data sharing agreements. By focusing on service quality rather than data quantity, this principle promotes fair competition and equal opportunities for all participants.

3. **Decentralised soft infrastructure**

   The data sharing infrastructure is not a monolithic centralised IT infrastructure. Instead, it is a “soft infrastructure” made up of agreements in all disciplines: functional, technical, operational, legal and economic, which allow data spaces to be interoperable with each other. This principle involves developing decentralised technologies and frameworks that support interoperability and data sharing without centralising control. There are functional and non-functional requirements, such as interoperability, portability, findability, security, privacy and trustworthiness, which all initiatives that join the European data space must satisfy.

4. **Public-private governance**

   Effective governance is required to ensure design, creation and maintenance of data spaces on an equitable basis. All stakeholders, including users (individuals, companies), data service providers, and their technological and professional partners, must feel represented and engaged. This principle involves establishing collaborative governance bodies that include representatives from both public authorities and private sector entities. These bodies are responsible for creating and enforcing rules, regulations, and policies that promote fair and transparent data management and sharing.

3.2 European data spaces architectures

3.2.1 IDS-RAM

In this chapter, we review the IDSA Reference Architecture Model (IDS-RAM 4.0) according to the methodology defined by PLIADES T2.1 coordinator.

To examine data space architectures, we can delve into the following aspects for analysis:

3.2.1.1 Conceptual foundations

The International Data Spaces (IDS) is a virtual data environment that leverages existing standards and technologies, as well as widely accepted governance models in the data economy, to facilitate secure and standardized data exchange and linkage within a trusted business ecosystem. This framework supports the creation of smart-service scenarios and innovative cross-company business processes, all while ensuring data sovereignty for data owners.

Goals of the International Data Spaces

Data sovereignty is a fundamental principle of the International Data Spaces.

> Data Sovereignty is the ability of a natural or legal person to exclusively and sovereignly decide concerning the usage of data as an economic asset [5].

It refers to the ability of individuals or corporate entities to maintain full control over their data. The IDS initiative provides a Reference Architecture Model to support this capability, addressing the requirements for secure and trusted data exchange within a business ecosystem.

3.2.2 Data Sovereignty as a Key Capability

The digital economy is characterized by two major developments:

-Data is becoming a strategic resource.

-Companies are increasingly collaborating within business ecosystems.

These trends create a fundamental conflict: while companies need to exchange data within these ecosystems, they also feel a heightened need to protect their data due to its growing importance. This conflict is intensified for companies deeply involved in multiple ecosystems, where the value of data significantly impacts collaborative success.

Data sovereignty addresses this conflict by balancing the need to protect data with the necessity of sharing it. It is a crucial capability for companies to thrive in the data economy. Achieving this balance involves closely examining the data, as different types of data require varying levels of protection and contribute differently to the value of collaborative efforts.

1. Strategic Requirements of the International Data Spaces

The IDS aims to meet several strategic requirements to facilitate a secure, trusted, and efficient data ecosystem:

- **Trust**: Trust is foundational to IDS. Each participant is rigorously evaluated and certified before gaining access to the trusted business ecosystem.
- **Security and Data Sovereignty**: IDS employs state-of-the-art security measures, ensuring that each technical component is evaluated and certified. Data sovereignty is maintained by allowing data owners to attach usage restriction information to their data. Data consumers must accept these usage policies to access the data.

- **Ecosystem of Data**: IDS promotes decentralized data storage, meaning data remains with its owner until transferred to a trusted party. Comprehensive descriptions of data sources, their value, and usability, along with domain-specific data vocabularies, are integrated. Metadata Brokers provide real-time data search services within the ecosystem.
- **Standardized Interoperability**: The IDS Connector, a central component of the architecture, is available in various implementations from different vendors. Despite this variety, all Connectors can communicate seamlessly with each other and other components within the ecosystem.
- **Value-Adding Apps**: IDS supports the injection of apps into the IDS Connectors to enhance data exchange processes. These apps provide services such as data processing, format alignment, and protocol management. Additionally, remote execution of data analytics algorithms is facilitated.
- **Data Markets**: IDS enables the creation of innovative, data-driven services and new business models. It provides mechanisms for clearing, billing, and creating domain-specific metadata broker solutions and marketplaces. Templates and methodological support help participants specify usage restrictions and legal information.

These strategic requirements ensure that the International Data Spaces provide a robust framework for secure and standardized data exchange, fostering trust and innovation within the digital economy.

2. Data driven business ecosystems

Ecosystems thrive on collaboration, where no single member can innovate alone; all members must contribute for mutual benefit, ideally achieving an equilibrium of shared advantages.

In a data-driven business ecosystem, data is the strategic resource that members use collectively to create innovative value offerings. Success hinges on sharing and jointly maintaining data, enabling partners to support end-to-end customer processes effectively.

International data spaces operate across all industries and domains of activity. Consequently, technologies, methods, organizational concepts, and governance structures must be designed generically to be universally applicable. At the same time, IDS concepts must be adaptable to key cross-sector domains with specific requirements.

The Data Spaces Radar

The Data Spaces Radar is a publicly accessible tool that provides a comprehensive overview of data space initiatives worldwide, offering insights into their sectors, locations, development stages, and use cases. Since its inception, the Radar has recorded over 140 entries.

To enhance its industry relevance, the International Data Spaces Association (IDSA) now hosts and maintains the tool, with additional contributions from the project community. The Data Spaces Support Centre (DSSC), a key part of the EU's Digital Europe program, coordinates EU-funded data space actions to ensure coherence, interoperability, and economies of scale through common practices and tools.

On the figure below we can see a screen shot of the Data Space Radar web interface that is available on the IDS website.

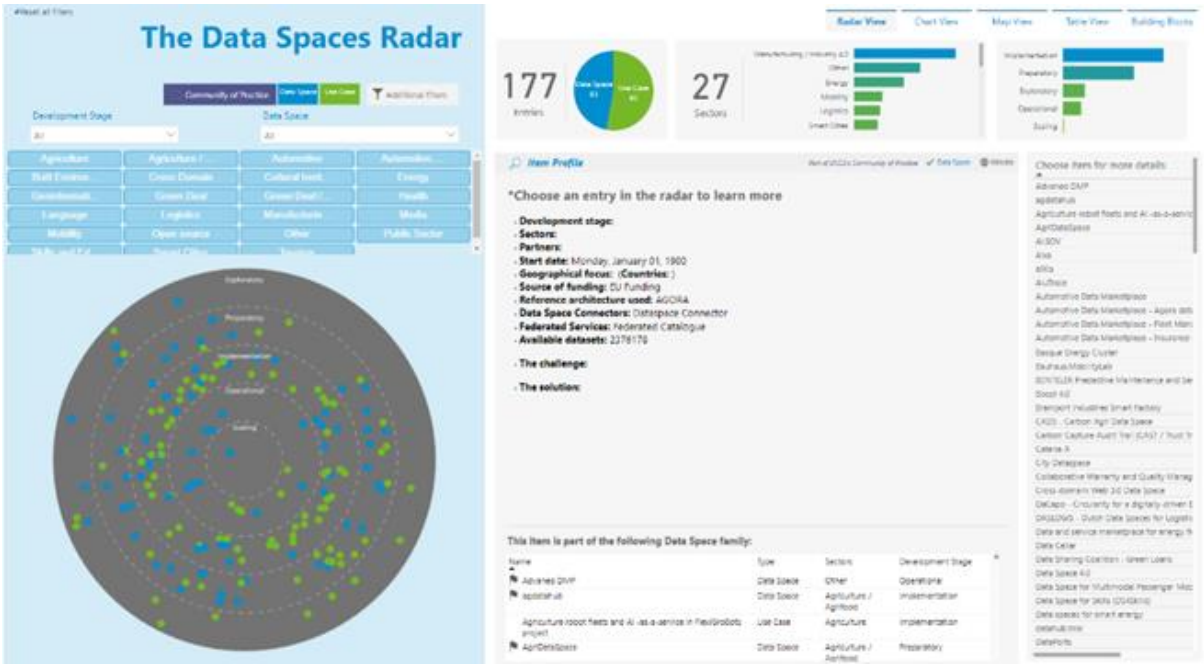

Figure 1. The Data Spaces Radar web interface, a publicly available tool providing an overview of the data spaces initiatives worldwide[1].

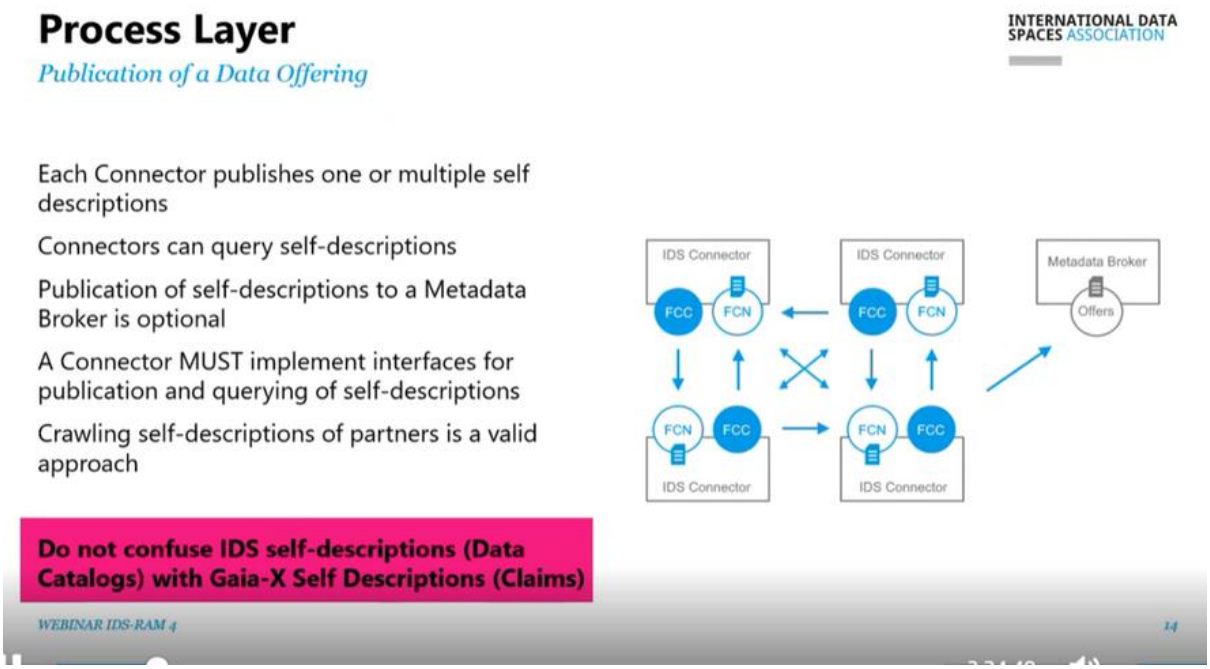

Figure 2. The process layer of IDSA architecture

3.2.2.1 The IDSA Information model

The ecosystem of the International Data Spaces (IDS) involves various tasks performed by participants as outlined in the IDS-Reference-Architecture Model. These tasks include managing relevant objects and activities throughout their lifecycle. Among these objects are vocabularies, which are ontologies,

[1] https://www.dataspaces-radar.org/

reference data models, or metadata elements used to describe datasets, usage policies, apps, services, and data sources.

A Vocabulary Intermediary manages and offers these vocabularies, assuming roles such as Vocabulary Publisher and Provider. Vocabularies are governed by standardization organizations and are crucial for annotating and describing data assets. These assets include the IDS Information Model, domain-specific vocabularies, and legal terms, all essential for the scalability and success of IDS.

There is no exclusive role for creating vocabularies; standardization organizations and industrial associations usually define them. Multiple vocabularies for the same context can exist, offering both standardization for compatibility and flexibility for competitiveness. Domain-specific adaptations, or Application Profiles, may be used to describe various IDS components, and independent domain-specific vocabularies may describe resource content and concepts.

The Vocabulary Hub in IDS manages vocabularies throughout their lifecycle, distinguishing between the Design Phase and the Runtime Phase from the perspectives of data providers and consumers.

Common semantic data models

The IDS Information Model based on DCAT and ODRL is the basic semantic model for IDS-based data spaces [78]. Each data space might have to enrich the Information Model with domain-specific information, which is not part of the Information Model. The Data Space Instance is responsible for 'standardizing' and development common semantic data models within the data space instance. The Data Space instance may make use of any mean for putting standards and developments in the ecosystem, as standardization through Standard Development Bodies (SDOs) is not always feasible and reasonable, an agreed structure in the ecosystem could also be considered as standard in this context.

More information about semantic interoperability: Position Paper Semantic Interoperability In Data-Spaces (internationaldataspaces.org) [10].

3.2.2.2 Governance mechanisms

The IDSA Rulebook provides a clear guideline for the mandatory and optional requirements of Data Spaces.

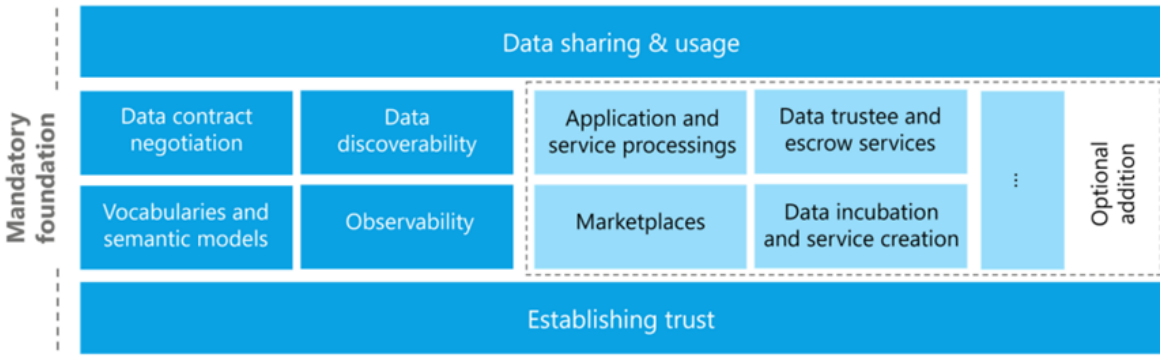

Figure 3. Foundational concepts of Data Spaces (source: IDSA Rulebook)

The Data Space Governance Authority (DSGA) is responsible for establishing the policies and rules of the data space. This role can be carried out by one entity, but also by multiple (or even all) participants. In a centralized data space, this could be the operating company. In a federated data space, this function would be performed by the federator(s) agreeing on the rules, while in a fully decentralized data space, various mechanisms are available to the participants. The mechanisms in a decentralized

data space enable participants to agree on the set of policies and their enforcement, thus sharing responsibility for the data space governance authority function.

To implement a DSGA and create a data space, three steps approach are needed:

1. Create an identity for the data space
2. Provide a self-description involving:
   - Membership policies
   - Trust anchors and trust frameworks
   - Attributes that will help participants decide which level of trust to apply for
   - Use of the technical components as required, according to the design
   - Participant registry
   - Registration service
     - Provide the workflow to apply for membership
     - Validate whether applicants comply with membership requirements
     - Issue membership credentials
     - Revoke membership credentials
3. Provide a discovery mechanism for the data space (website, contact form, etc.)

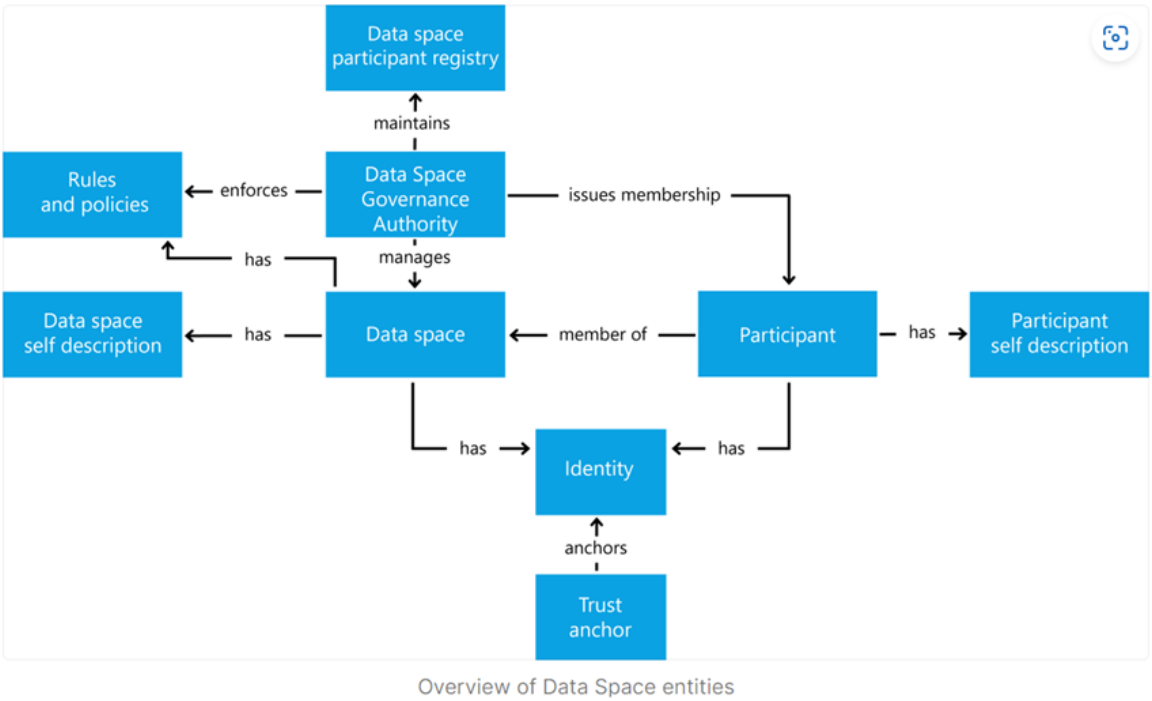

Figure 4. Conceptual entities of a Data Space defined by the IDSA Rulebook

3.2.1.4 Security features

The International Data Spaces (IDS) Reference Architecture Model (RAM) [78] emphasizes security as a cornerstone for ensuring trust among participants who exchange and share data. Here is a summary of the security perspective outlined in the IDS RAM[75]:

Strategic Security Requirements

- **Secure Data Supply Chains**: Essential for building and maintaining trust among participants.
- **Security Architecture**: Focuses on device identification, secure communication, data exchange, and post-exchange data usage control.

IDS Connectors

- **Implementation**: Ensures the practical application of security specifications in daily interactions within the IDS.
- **Layered Security Approach**: Security requirements are detailed for different layers of an IDS connector.

Identity and Trust Management

- **Decentralized Approaches**: Utilizes decentralized trust frameworks to manage identities and ensure trust among devices and entities within the IDS.

Security Measures on Different Layers

1. **Platform Layer**:
    a. **System Security**: Ensures confidentiality and integrity through a Trusted Computing Base (TCB) consisting of critical hardware and software components.
    b. **Deployment Scenarios**: TCB requirements vary depending on deployment scenarios.
2. **Application Layer**:
    a. **Secure Execution**: Requires a secure platform for the isolated execution of applications.
    b. **Security Mechanisms**: Applications must integrate with platform security mechanisms and fulfil specific security requirements.
    c. **Application Authenticity**: The platform verifies application authenticity and integrity through signature checks on App Manifests.
    d. **Usage Control Policies**: Enforces licensing policies, such as usage time limits or instance restrictions.

Communication Security

- **Secure Data Transfer**: Ensures secure communication between IDS components by:
    - Identifying, authenticating, and authorizing components.
    - Protecting data confidentiality and integrity.
    - Establishing and using secure communication channels.
    - Continuous Dynamic Trust Monitoring of IDS components.

Usage Control

- **Access Control Models**: Implements various models like RBAC and ABAC to restrict and authorize access to resources.
- **XACML Standard**: Uses this standard to define access control policies with components like subject, action, resource, and environment.

The IDS RAM's security perspective provides comprehensive measures to ensure secure data exchange [8] and processing within the IDS ecosystem, focusing on trusted identities, secure communication, robust platform and application security, and effective access control mechanisms.

3.2.1.5 Standardization

IDSA is committed on contributing to standardization activities thanks to the commitment of IDSA the dedicated efforts of IDSA members and technical staff who actively engage in standards committees. By both observing and leading these committees, IDSA strategic partnerships with key standardization bodies are established and advising the European Commission on advancing global standards for data spaces. Utilizing assets such as the IDS Reference Architecture Model (IDS RAM), the IDS Rulebook, and the Dataspace Protocol, we contribute to developing crucial technical specifications for implementing data spaces.

European Standardization Initiatives

In the European context, regulations such as the Digital Markets Act (DMA), Data Governance Act (DGA), Data Act, and AI Act underscore the need for specific standardization efforts. Collaborative research with CEN/CENELEC is vital for addressing these standardization needs and enhancing Europe's digital economy. Initiatives like the CEN/CENELEC Focus Group on "Data, Dataspaces, Cloud, and Edge" and the proposed Technical Committee on Data Management, Data Spaces, Cloud, and Edge Computing are key to shaping European standardization policies.

Collaboration with the European Commission

IDSA works closely with the European Commission to ensure that standardization efforts align with EU policy priorities. The Data Spaces Support Centre (DSSC) collaborates with the European Data Innovation Board to recommend guidelines for unified European data spaces. By improving data interoperability and ensuring compliance with data protection regulations, these efforts significantly contribute to the EU's digital transformation agenda.

The IDSA's community dedication to driving global standardization in data spaces is paving the way for a more connected, efficient, and innovative future for businesses and societies worldwide.

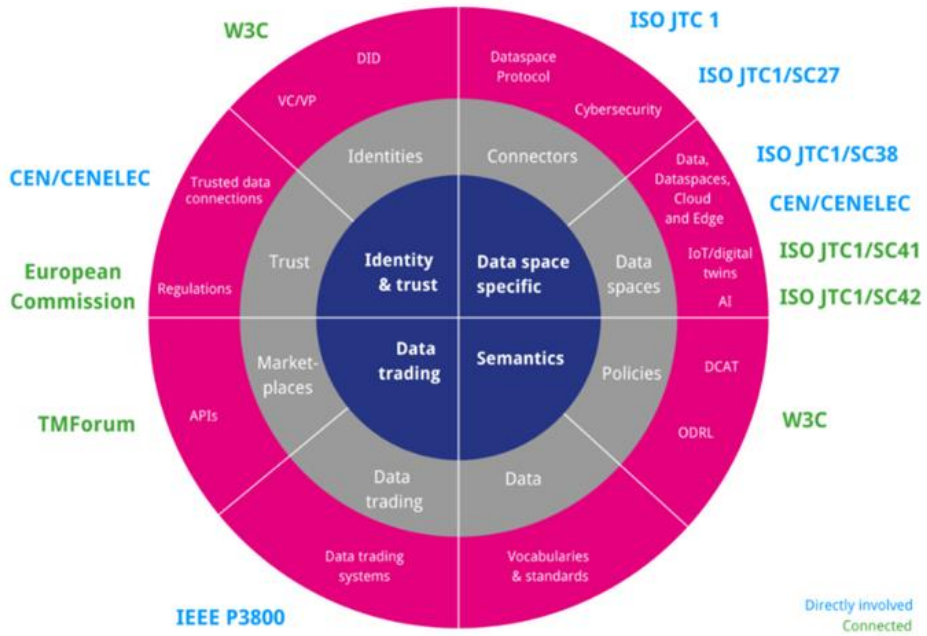

Figure 5. Complete overview of data spaces standardization landscape and committees[2].

IDSA is actively participating on the following standardisation bodies:

[2] https://internationaldataspaces.org/why/international-standards/

- CEN/CENELEC Trusted Data Transactions,
- CEN/CENELEC Focus Group Data, Data spaces, Cloud
- ISO/UIEC JTC1/SC 38 Cloud computing and distributed platforms

3.2.1.6 Added services such as data anonymization tools, data quality management functionalities

IDSA Reference Architecture Model provides anonymization when the Data Space participants use different Data Services. The first layer of anonymization is the data access control phase, which enforces a security policy for the participants. Specifically, only accepted Connectors equipped with security protocols are granted access.

After access control is conducted, data usage control follows. Smart contracts are used to define the data that are secure when shared with the data consumers. Another way anonymization is achieved is by offering data only in an aggregated form to ensure the users are anonymized. Also, sensitive data is replaced by adequate substitutes that do not affect the homogeneity of the data, while anonymity is accomplished.

Another core component of added services is extensions handling data quality issues. Prometheus-X, an IDSA partner, implements an AI data enrichment strategy, based on Automatic Natural Language Processing models, to improve the interoperability of the ecosystem while increasing data accuracy and making them more relevant to data consumers.

3.2.1.7 Does the architecture offer functionalities for auditing data access and lineage tracking to ensure accountability and compliance with regulations?

This functionality in the IDS RAM5.0 is called observability:

The observability function in regulated data spaces ensures that data sharing processes are transparent and accountable, mainly for legal compliance and business operations. This function is critical for proving that only authorized entities have processed data and for enabling marketplace and billing activities via a trusted third party.

There are different architectural approaches to implement observability:

1. **Centralized architecture**: In this model, a central observer like a clearing house or auditor monitors the data transactions. However, this approach has notable drawbacks:
    a. It creates a single point of vulnerability that could compromise the sharing of mission-critical data.
    b. The central observer accumulates valuable data on all data-sharing activities (DCAs), making it a potential target for exploitation and attacks.
2. **Federated architecture**: This model suggests distributing the observatory functions among multiple observers to mitigate the risks associated with a centralized system. It helps spread the information load and reduces the chance of errors and misuse.
3. **Decentralized architecture**: This is the most robust model, where each participant maintains their own logs about the DCAs, thus maintaining at least two copies of each log linked by a correlation ID. This structure minimizes risks by:
    a. Ensuring no single point of failure or attack.
    b. Facilitating independent verification of logs by matching entries from different participants.

c. Reporting irregularities directly to involved parties or regulators as needed.

Additionally, the system allows for trusted third parties, such as industry auditors, to act as observers. These auditors, validated by a governmental trust anchor, can request log data which is then shared under strict access policies. This mechanism ensures that only observers with proper credentials and bound by data-sharing contracts can access sensitive information.

Moreover, the system mandates automatic logging of observer actions, thus supporting a trust framework where even auditors can be audited. The design also proposes that audit data be made available as events or streams by default, simplifying access for trusted auditors who can negotiate relevant contracts directly.

Optional roles like payment clearance, notary services, and regulatory reporting can also be integrated into the system, further enhancing its functionality and compliance capabilities.

The IDS RAM 4.0 Clearing House.

The IDS Clearing House consists of an IDS Connector and bases all its functions on a logging service that records information relevant for clearing and billing as well as usage control. The information sent to the Clearing House is defined in the Process Layer of the ISD RAM4.0 [78].

The Clearing House uses this information to provide a Clearing and Settlement Service on the basis of usage contracts and helps with the automation of payments between Data Provider and Data Consumer. It can also use this information to provide a Billing Service to allow the Data Space Operator the billing of the participants. The UC Claim Validation service uses the logged usage control data to allow the validation of usage claims on resources.

3.2.1.8 Does the architecture address the data sovereignty?

Data sovereignty is a core principle of the International Data Spaces (IDS), the IDSA RAM4.2 refer on the data usage policies & usage enforcement hat, in the IDS architecture, the Data Owners and Data Providers can always be sure their data is handled by a Data Consumer who fulfils the usage policies specified. Each participant can define usage policies and attach them to outbound data. Policies might include restrictions, such as disallowing persistence of data, or disallowing transfer of data to other parties, for example. More information about usage policies and usage enforcement can be found the IDS rulebook, where it is emphasized that digital sovereignty starts with identity control, which is crucial for secure and trusted data exchange within a data space. A federated identity system enables participants to exert control over their data, choosing what to share, with whom, and under what conditions. This system allows for distributed control without relying on a single central authority.

The governance of a data space, managed by a Data Space Governance Authority (DSGA) mentioned in 3.2.1.3, is responsible for establishing the policies and rules. This authority can vary from a centralized entity to multiple federators or even a fully decentralized model, depending on the data space's structure. Policies are crucial for maintaining trust and ensuring proper data usage. They can include:

- **Membership Policies**: Define the requirements for participants to join a data space, ensuring only verified entities with appropriate attributes can participate.
- **Access Policies**: Control access to data contracts based on participant attributes, allowing or restricting visibility of data offers.
- **Contract Policies**: Specify terms and conditions for data contracts, including technical and legal attributes required for negotiation and compliance.
- **Usage Policies**: Dictate how data can be used after transmission, considering the data's value, sensitivity, and applicable regulations.

Policies can express prohibitions, obligations, and permissions, and they may vary in complexity. Effective management of these policies ensures that data is shared securely and responsibly, supporting data sovereignty while enabling collaboration across different data spaces.

3.2.1.9 Does the architecture address inter- and intra-data spaces interoperability?

The IDSA architecture facilitates both inter- and intra-data space interoperability through a structured set of principles, protocols, and policies, ensuring seamless data exchange and collaboration across and within data spaces.

Interoperability Models:

- **Intra-Data Space Interoperability**: Involves participants interacting within a single data space under the governance and protocols defined by the DSGA. Participants must adhere to established identity protocols, trust frameworks, and semantic models.
- **Cross-Data Space Interoperability**: Requires participants to access and exchange data between different data spaces. Participants must be members of both spaces, supporting the necessary protocols and semantic models. Alternatively, DSGAs and legal entities from different data spaces can collaborate to reduce the complexity for participants, aligning protocols and semantic models to facilitate smoother cross-data space interactions.

Intra-Data Space Interoperability:

This concerns the ability of different components within a single data space to operate together effectively. IDSA addresses this through:

Technical Interoperability: Access rights and usage control mechanisms ensure that data exchanges within the data space are secure and follow standardized procedures, facilitating smooth interactions between participants.

- **Data Sovereignt**y: Even within a single data space, data owners maintain control over their data, specifying who can access it, for what purpose, and under what conditions, thereby upholding data sovereignty.

Cross-Data Space Interoperability:

This refers to a scenario where a participant needs to access data from multiple, distinct data spaces. The IDSA addresses this through:

- **Dataspace Protocol**: Adoption of the Dataspace Protocol across different data spaces ensures that all services and protocols are aligned and compatible, allowing participants to access and exchange data seamlessly.
- **Semantic Interoperability**: By adhering to the common semantic model provided by IDSA, participants across different data spaces can interpret data consistently. This reduces the risk of misalignment and miscommunication, enabling high levels of interoperability.

Overall, the IDSA framework provides a comprehensive approach to ensure interoperability, enabling data to be shared and reused effectively while maintaining control and compliance with relevant regulations.

3.2.1.10 How well can the architecture scale to handle increasing data volumes and user demands while maintaining efficient performance?

The International Data Spaces Reference Architecture Model (IDS RAM) is designed to facilitate secure and standardized data exchange among various parties. A key aspect of its architecture is the separation of control and data planes, which contributes significantly to its scalability. Let's explore how this design aids scalability and maintains performance despite growing data volumes and user demands.

Scalability through separation of Control and Data Planes

Data space connector should implement both a "Control Plane" and a "Data Plane". The Data Plane performs the actual transfer of data from the data provider to the data consumer. Different Data Planes support different data transfer or communication protocols, such as HTTPS, S3 file transfer, REST, or messaging queues.

1. Independent scaling of planes: The separation of control and data planes in the IDS RAM allows each plane to be scaled independently based on specific needs. This is crucial because the control plane, which handles decisions and permissions, might not require the same scalability solutions as the data plane, which deals with the actual data transfer.

2. Flexibility in data handling: With this architecture, different data transfer protocols can be implemented without necessitating changes to the control mechanisms. This flexibility means that as data volumes grow or as different types of data are integrated into the system, the data plane can adapt without affecting the overall control structure. This adaptability is vital for handling large-scale data operations efficiently.

3. Separate decision-making from action-taking: By decoupling decision-making (control plane) from action-taking (data plane), the IDS RAM ensures that enhancements or modifications in the process flow (like introducing new algorithms for data handling or new security measures) can occur in one plane without disrupting the other. This separation aids in maintaining system integrity and performance even as user demands evolve.

4. Enhanced performance management: The ability to independently manage and optimize each plane leads to better performance management. For instance, if the data plane requires more resources due to an increase in data throughput, it can be scaled up by adding more bandwidth or servers specifically targeted to data transfer without overloading the control plane.

Overall, the IDS RAM's architectural design, with its clear distinction between control and data planes, not only enhances scalability but also ensures that performance is optimized across different aspects of the data exchange system. This model provides a robust framework that can handle increasing volumes of data and a growing user base effectively, making it a resilient choice for organizations aiming to manage their data exchange ecosystems efficiently.

3.2.1.11Adherence to relevant regulations, e.g., GDPR.

IDSA has contributed to the discussion about the Data Governance Act (DGA) regulations on “data intermediation services” and the role of data intermediaries in data spaces [11]. See the document on this link: IDSA Position Paper | Reflections on the DGA and Data Intermediaries (internationaldataspaces.org)

https://internationaldataspaces.org/wp-content/uploads/Reflections-on-the-DGA-and-Data-Intermediaries.pdfThe IDS rule book addresses the legal dimension of data spaces, gives an overview of the regulatory framework and describes IDSA's approach of compliance with regulatory requirements and contractual agreements.

The IDS rule book discusses the evolving regulatory and legal landscape surrounding data governance within the EU. A summary of the key points is provided:

1. Regulatory Framework Overview:
    - The text highlights the fragmented nature of data regulation due to partial application of intellectual property rights, trade secret protection, and personal data protection laws. To address this, the EU Commission has introduced several strategies and legislative measures, including the "European strategy for data," aiming to create a unified European data space.
2. Key Legislative Acts:
    - Digital Markets Act (DMA) and Digital Services Act (DSA): Focus on harmonizing rules for data governance, access, and use.
    - Data Governance Act (DGA): Enacted to enhance public access to protected public sector data and promote data sharing for altruistic purposes. It focuses on ensuring that public sector data, which is legally protected, is made more available for innovation.
    - Data Act Proposal (DA-E): Proposed to ensure fair access to and use of data, addressing the needs of SMEs and start-ups by establishing a clear contractual framework for data access and sharing.
3. Operationalization of Data Governance:
    - A comprehensive approach is needed to navigate the existing regulatory patchwork and implement new EU legislative agendas. This includes developing a four-pillar data governance framework covering substantive rights, contractual dimensions, organizational aspects, and technical implementation.
4. Legal Agreements and SITRA Rulebook:
    - Discusses the gaps in current legal frameworks for data transactions, which do not fully address the needs of the data economy. The SITRA rulebook, updated periodically, offers a model for data sharing networks, providing legal, business, technical, security, and ethical guidelines.
    - IDS (International Data Spaces) Rulebook: Aligns with SITRA's principles, focusing on data sovereignty and trust. IDS utilizes SITRA’s rulebook as a foundation but proposes modifications to suit specific data sharing contexts.
5. Future Legal Developments:
    - Continuous monitoring and updating of legal frameworks are essential to ensure compliance with evolving data governance requirements. IDSA (International Data Spaces Association) is actively involved in these developments through a legal framework task force.

In summary, the text outlines the complex and multi-dimensional efforts underway in the EU to create a cohesive and fair framework for data governance and sharing, addressing both current gaps and future needs in the digital economy.

GDPR is applied at use case level not at the architecture level.

3.2.1.12 Protocols used for communication.

The **Dataspace Protocol** is a set of specifications designed to facilitate interoperable data sharing between entities governed by usage control and based on Web technologies. **This specification does not cover the data transfer process as such.**

While the data transfer is controlled by the **Transfer Process Protocol** mentioned above, e.g., the initiation of the transfer channels or their decommissioning, the data transfer itself and especially the handling of technical exceptions is an obligation to the Transport Protocol.

As an implication, the data transfer can be conducted in a separated process if required, as long as this process is to the specified extend controlled by the **Transfer Process Protocol**.

Nevertheless, illustrative message examples are provided in the **Transfer Process Protocol section**. [12]. The best practices section may contain further non-normative examples and explanations.

The dataspace protocol specifies how to agree on the data transfer types. Dataset [13] transfers are characterized as push or pull transfers and its data is either finite or non-finite. This section describes the difference between these types.

Push Transfer

A push transfer is when the Provider's [14] data plane initiates sending data to a Consumer [15] endpoint. For example, after the Consumer has issued a Transfer Request Message, the Provider begins data transmission to an endpoint specified by the Consumer using an agreed-upon wire protocol.

Pull Transfer

A pull transfer is when the Consumer's data plane initiates retrieval of data from a Provider endpoint. For example, after the Provider has issued a Transfer Start Message, the Consumer can request the data from the Provider-specified endpoint.

3.2.1.13 Current stage of development.

There are hundreds of reference implementations based on IDS RAM4.0 and at this stage of development there are some few implementations based on DID and the Dataspace Protocol that is the base for the future IDS RAM5.0

A good example is the Catena X data space. So, mature enough for implementing real data spaces. Also, Eclipse Dataspace Components (EDC) provides a comprehensive framework that users can implement and customize connectors based on specific requirements.

3.2.1.14 FOR ADDITIONAL INFORMATION:

Link to the recording of the webinar:

https://n-cloud1.robotics.iti.gr/index.php/s/4ifFwqforZNLo3N

3.2.3 GAIA-X

3.2.3.1 Conceptual foundation

Source: Architecture document - Gaia-x - DRAFT version 1702083 [16]

1. **Goals** (e.g., secure data exchange in a specific industry): Gaia-X aims to create a **federated and open data infrastructure** based on European values regarding **data sovereignty** and the cloud. Its mission is to design and implement a data exchange architecture with common standards, best practices, tools, and governance mechanisms.
2. **Target use cases** (generic, adaptable to various domains, specific domain): It is focused on enabling Data Ecosystems and Infrastructure, using elements explained in the Gaia-X Conceptual Model, the Operating Model, and Federation Services along with the Gaia-X Trust Framework. These ecosystems are adaptable to various domains, such as automotive (Catena-X) or agriculture (Agdatahub).

3. **Data model focus** (on a specific data model (structured, unstructured), or flexible for handling different data types)**:** The Gaia-X Conceptual Model describes all the concepts within the scope of Gaia-X and the relationships between them. It focuses on interoperability and portability of resources within and across Gaia-X-based ecosystems and provides data sovereignty in a distributed ecosystem environment. It is flexible to handle different types of data, including physical and virtual resources.

3.2.3.2 Governance mechanisms

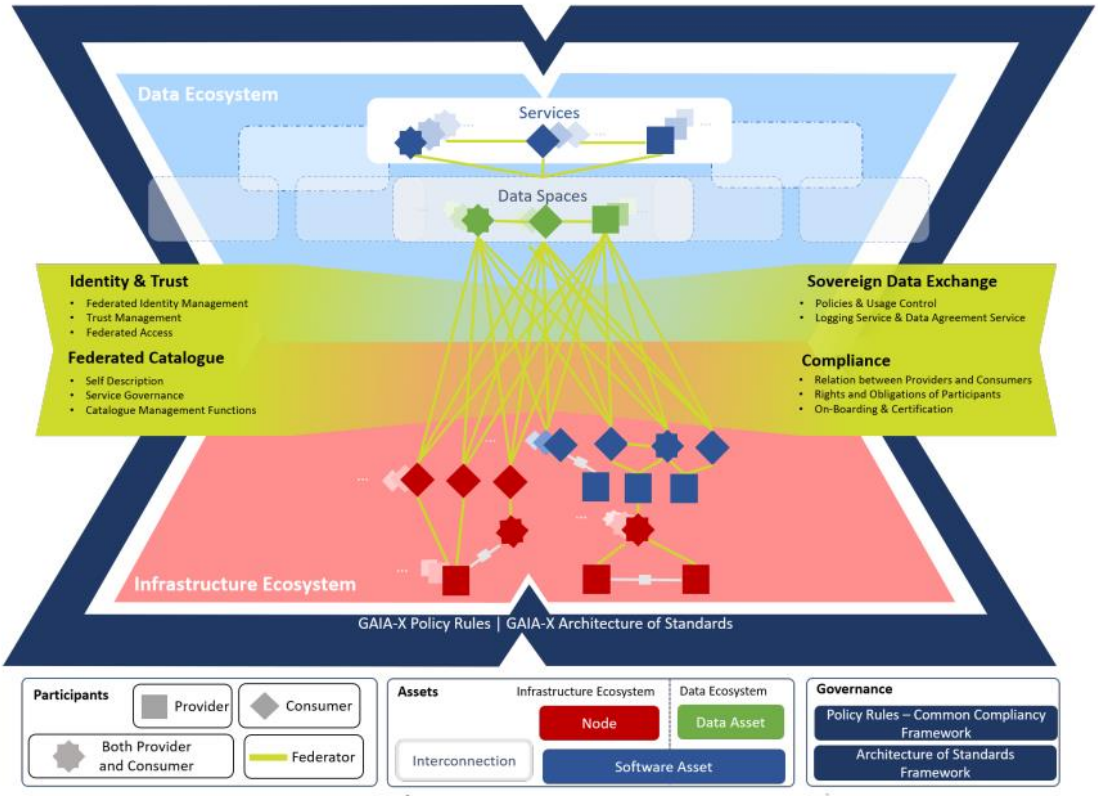

Figure 6. Gaia-X Ecosystem Visualization [3].

- **Participant Roles**: Gaia-X defines clear roles for participants, including Providers, Consumers, and Federators, facilitating interaction and governance within the ecosystem.
- **Federation Services**: These services establish interoperability and portability of resources, ensuring trust among participants and facilitating sovereign data exchange.
- **Trust Framework**: Gaia-X AISBL (The Gaia-X European Association for Data and Cloud) defines a trust framework manifested in services such as the Gaia-X Registry and the Gaia-X Compliance Service, supporting digital governance.
- **Gaia-X Registry**: This is a central database that stores information about data and service providers within the Gaia-X network. It serves as a directory where participants can register their offerings and capabilities, making it easier for others to discover and connect with them. The registry promotes transparency and facilitates the efficient utilization of resources within the Gaia-X ecosystem.
- **Gaia-X Compliance Service**: This service ensures that participants adhere to the principles, standards, and regulations set forth by Gaia-X. It provides tools and mechanisms for assessing

[3] https://docs.gaia-x.eu/technical-committee/architecture-document/22.10/

and verifying compliance with Gaia-X requirements, such as data sovereignty, security, and interoperability. The compliance service helps maintain trust and reliability within the ecosystem by ensuring that all participants operate in accordance with Gaia-X guidelines.

- **Gaia-X Labels**: These provide an optional scheme for Gaia-X compliance and support business, or regulatory decisions based on attributes compatible with Gaia-X.

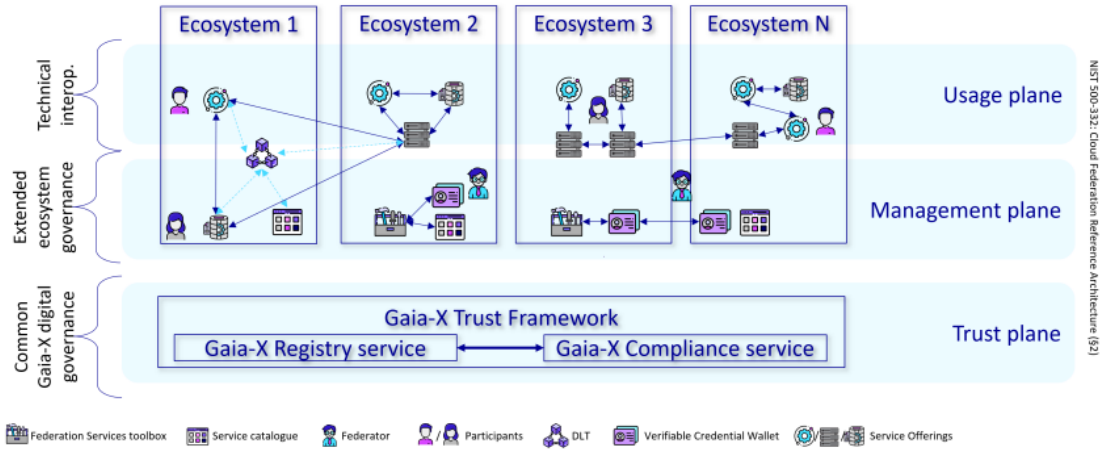

Figure 7. Gaia-X planes that represent the three levels of interoperability [4].

3.2.3.3 Security features

Since Gaia-X is based on distributed data ecosystems, it recommends a few security features that ecosystem managers should implement to ensure data reliability.

Data Sovereignty: Gaia-X emphasizes the sovereignty of data, ensuring that data owners have control over how their data is stored, processed, and shared. This principle helps prevent unauthorized access and usage of data by ensuring that it remains under the jurisdiction of its owner.

- **Encryption**: Gaia-X promotes the use of encryption techniques to secure data both in transit and at rest. This ensures that even if data is intercepted or compromised, it remains unintelligible to unauthorized parties.
- **Identity and Access Management (IAM)**: IAM systems control and manage access to resources within the Gaia-X ecosystem. By enforcing strict authentication and authorization mechanisms, IAM helps prevent unauthorized users from accessing sensitive data or resources.
- **Audit Trails and Logging**: Gaia-X encourages the implementation of comprehensive audit trails and logging mechanisms to track user activities and system events. This enables the detection of suspicious behavior, compliance monitoring, and forensic analysis in the event of security incidents.
- **Compliance Framework**: Gaia-X establishes a compliance framework that defines standards, guidelines, and best practices for security within the ecosystem. Compliance with these standards ensures that participants adhere to security principles and regulations, enhancing overall trust and confidence.

[4] https://docs.gaia-x.eu/technical-committee/architecture-document/22.10/

- **Secure Communication Protocols**: Gaia-X promotes the use of secure communication protocols, such as TLS/SSL, to encrypt data transmission between services and endpoints. This prevents eavesdropping and tampering with data during transit.
- **Cybersecurity Measures**: Gaia-X encourages the implementation of robust cybersecurity measures, including intrusion detection systems, firewalls, and endpoint security solutions, to protect against various threats such as malware, phishing, and DDoS attacks.

3.2.3.4 Standardization

- **Contractual Framework**: A legally binding agreement between providers and clients is required, which may be a contract with documentation accessible to both parties.
- **General Material Requirements**: These include provisions regarding service interruptions, business continuity, service and data usage rights, and service changes.
- **Data Protection**: GDPR requirements must be met when processing personal data, defining roles, responsibilities, and technical and organizational measures.
- **Cybersecurity**: Information security measures must be implemented, and risks associated with information security managed.
- **Portability**: Practices for changing providers and transferring data in structured and commonly used formats must be facilitated.

3.2.3.5 Does the architecture offer functionalities for auditing data access and lineage tracking to ensure accountability and compliance with regulations?

These features are crucial for maintaining transparency, demonstrating compliance with data protection laws such as GDPR, and enabling effective governance within the Gaia-X ecosystem.

- **Auditing Data Access**: Gaia-X facilitates the logging and monitoring of data access activities, allowing organizations to track who accessed which data, when, and for what purpose. This audit trail helps enforce access controls, detect unauthorized access or misuse of data, and demonstrate compliance with regulatory requirements.
- **Lineage Tracking**: Gaia-X also supports lineage tracking, which involves tracing the origins, transformations, and movements of data throughout its lifecycle. By capturing metadata and lineage information, organizations can establish data provenance, understand data dependencies, and ensure data quality and integrity. Lineage tracking is particularly important for compliance with regulations that mandate data governance and accountability, as it enables organizations to demonstrate the lineage and authenticity of data used for decision-making and reporting purposes.

3.2.3.6 Does the architecture address the data sovereignty?

Gaia-X addresses data sovereignty by creating a federated open data infrastructure based on European values. It designs and implements common standards, best practices, tools, and governance mechanisms for data sharing, enabling trusted and secure exchanges of data across multiple actors.

3.2.3.7 Does the architecture address inter- and intra- data spaces interoperability?

Gaia-X aims to address both inter- and intra-data space interoperability:

Inter-Data Space Interoperability:

- Gaia-X facilitates interoperability between different data spaces or ecosystems, which can be industry-specific, regional, or organizational.
- It provides a framework for data sharing, exchange, and collaboration across these diverse spaces.
- By defining common standards, protocols, and interfaces, Gaia-X enables seamless communication between various data ecosystems.

Intra-Data Space Interoperability:

- Within a single data space (e.g., a specific industry sector), Gaia-X ensures interoperability among different stakeholders, including data providers, consumers, and service providers.
- Gaia-X promotes transparency, trust, and secure data sharing within these individual data ecosystems.

In summary, Gaia-X's architecture is designed to foster interoperability at both the macro level (between different data spaces) and the micro level (within a specific data space).

3.2.3.8 How well can the architecture scale to handle increasing data volumes and user demands while maintaining efficient performance?

Gaia-X is designed to scale effectively to handle increasing data volumes and user demands while maintaining efficient performance. Several factors contribute to its scalability:

- **Federated Architecture**: Gaia-X utilizes a federated architecture, which allows for distributed data management across multiple nodes or providers. This decentralized approach enables horizontal scaling, where additional resources can be added as needed to accommodate growing data volumes and user demands.
- **Interoperability**: Gaia-X promotes interoperability among different data sources, services, and applications. By adhering to common standards and protocols, Gaia-X facilitates seamless integration and communication between disparate systems, allowing for scalability without compromising interoperability.
- **Distributed Computing**: Gaia-X supports distributed computing paradigms, such as edge computing and distributed processing, which distribute computational tasks across multiple nodes or devices. By harnessing the computational power of distributed infrastructure, Gaia-X can efficiently process large datasets and handle complex analytics workloads.
- **Scalable Infrastructure**: Gaia-X encourages the use of scalable infrastructure components, such as cloud services and containerization technologies, which can easily scale to accommodate growing demands. By leveraging scalable infrastructure, Gaia-X can seamlessly expand its capacity to meet the needs of users and applications.

3.2.3.9 Adherence to relevant regulations, e.g., GDPR

Source: Gaia-X_Policy-Rules_Document_v22.04_Final.pdf [17]

The GDPR [76] **only applies** in the case of processing **personal data**. By principle, this shall only apply to personal data that are processed and are subject to the commercial relationship between the customer and the provider, but not those personal data that are processed by the provider to establish and maintain such commercial relationship for its own purposes, e.g., contract handling and invoicing.

3.2.3.10 Protocols used for communication

Gaia-X does not specify mandatory communication protocols that connectors must use. Instead, it focuses on interoperability and flexibility. Connectors can use different protocols according to their needs and technical capabilities. This allows adaptation to various environments and existing systems.

3.2.3.11 Current stage of development

Gaia-X is still active, developing and improving tools that ensure data security and federation. One of the most notable milestones recently was the launch of GXDCH (Gaia-X Digital Clearing House) in March 2023.

The GXDCH is a node of verification of the Gaia-X rules, it is the go-to place to obtain Gaia-X compliance and become part of the Gaia-X ecosystem. Are non-exclusive, interchangeable multiple nodes operated by market operators, acting as a Gaia-X Federator. They operate and run services of the Gaia-X Framework (compulsory and optional), necessary to achieve compliance and support the onboarding of any Gaia-X adopter:

- Mandatory GXDCH components:
  - Gaia-X Registry
  - Gaia-X Compliance
  - Gaia-X notarisation service for the registration number
- Optional GXDCH components
  - Wizard
  - Wallet
  - Catalogue

The development has been in the testing phase since March 2024, and it is currently possible to create a Gaia-X compatible environment using one of the various enabled Hubs.[18]

3.2.4 FIWARE

3.2.4.1 Conceptual foundation

FIWARE was established to develop a robust, open ecosystem based on a public, open-sourced software platform standard [9]. This initiative facilitates the development of smart solutions, helping organizations transition into more intelligent operations. Technically, FIWARE offers a comprehensive collection of open-source software components. These components can be integrated and supplemented with other third-party platform elements to create systems that simplify the creation of data spaces and, more generally, smart digital solutions in various sectors, including cities, manufacturing, utilities, agrifood, and more. This approach not only enhances organizational efficiency but also fosters innovation in numerous application domains.

Initiatives such as FIWARE enable the fast road to market of data space prototypes and the extension of specific system components that can be built on top of FIWARE blocks.

1. **Goals** (e.g., secure data exchange in a specific industry): By creating open standards for the IoT ecosystem, FIWARE aims to facilitate interoperability between devices, platforms, and applications from different vendors. Moreover, FIWARE aims to accelerate innovation in the IoT and smart solutions domain.
2. **Target use cases** (generic, adaptable to various domains, specific domain): FIWARE has a strong focus on smart city applications. Also, the core functionalities and components of the platform are designed to be adaptable to various domains.

3. **Data model focus** (on a specific data model (structured, unstructured), or flexible for handling different data types): FIWARE leans towards a flexible approach for handling different data types, rather than strictly enforcing a single data model.

3.2.4.2 Governance mechanisms

Governance (Blueprint, DBSA Technical convergence) refers to the set of policies, rules, and standards that manage how data is accessed, shared, and used within the data space. Governance involves the structures and processes to ensure data security, privacy, interoperability, and compliance with legal frameworks, particularly around GDPR, the Data Act, or the AI Act. Governance in a data space requires the adoption of a number of businesses, operational, and organizational agreements across the actors participating.

- Business: Terms and conditions of the data sharing. Legal framework supporting the contracts.
- Operational: Policies that must be enforced during any operation (GDPR). Tools and global services that enable auditing or certain processes or the adoption of cybersecurity practices.
- Organizational: Bodies in charge of identifying product specifications, technology building blocks in a data space, and their requirements.

While the governance initiative related to the implementation and design of data spaces go beyond the technical requirements, FIWARE has already joined forces with other organizations like TM Forum or IUDX to support an open governance model following best open-source practices. Specifically, these efforts have been implemented in the context of the Smart Data Models Initiative that provided a library of Data Models based on JSON-LD, compatible with NGSI APIs and other RESTful interfaces Open API compliant.

Additionally, FIWARE works seamlessly with the architecture elements that the International Data Space Association (IDSA) is developing to create data spaces with trust and data sovereignty. FIWARE components have proven to integrate smoothly with other relevant building blocks in the context of data spaces.

3.2.4.3 Security features

FIWARE adopts the IDS Reference Architecture Model (RAM), implementing organization-level identity and access management. This includes using a Certification Authority and a Dynamic Attribute Provisioning Service, which, along with IDS Connectors, confirm participant identities and enforce security policies for data sharing permissions. FIWARE enhances security by facilitating user-level access control. This is managed through the Keyrock Identity Manager, which handles user identification, authentication, and authorization using standards such as OAuth2, OpenID Connect, and SAML 2.0 [19].

FIWARE employs the NGSI-LD API for secure data interchange, which ensures structured data exchange across systems. It also uses secure communication protocols like HTTPS and WSS for encrypted, real-time data exchanges. Moreover, data transactions within FIWARE can be securely logged using blockchain or other distributed ledger technologies to ensure traceability. The modular nature of FIWARE allows for the creation of tailored security solutions. Organizations can customize security policies to meet specific needs by leveraging FIWARE's flexible building blocks.

3.2.4.4 Standardization

FIWARE approaches standardization at all levels, from technical components and open-source initiatives to data transmission protocols and data models. FIWARE adopts the NGSI-LD API, which is a standardized API developed by the European Telecommunications Standards Institute (ETSI). This

API facilitates managing and exchanging context information across different systems, enabling seamless communication [20].

FIWARE's engagement with standardization bodies like ETSI is part of a broader collaboration that extends to other initiatives, such as the Smart Data Models Initiative. This initiative focuses on developing standardized data models essential for ensuring data consistency across different applications. These models are designed to be directly compatible with the NGSI-LD API and other RESTful interfaces to simplify integration and adoption. FIWARE's approaches to standardization are extensively community-driven, engaging a broad ecosystem of developers and researchers to participate in the refinement of standards. Each of the standards offered is available in the FIWARE catalog.

3.2.4.5 Added services such as data anonymization tools, data quality management functionalities

FIWARE architecture does not directly handles data manipulation tools such as anonymization or quality management. However, FIWARE supports data quality from a modelling perspective. Smart Data Models and its standardized format allow for a uniform data structure and semantics to ensure data quality. This supports the modelling of data validation processes that check data for accuracy, completeness and any other relevant metric. Similarly, with the components that support real-time processing and monitoring (i.e., CoatRack for API management), FIWARE allows for detecting anomalies, inconsistencies or outdated information.

3.2.4.6 Does the architecture offer functionalities for auditing data access and lineage tracking to ensure accountability and compliance with regulations?

FIWARE's Generic Enablers, like the Orion Context Broker, offer functionalities for access control and data authorization. Also, some Generic Enablers offer functionalities for logging user activities and data modifications. These logs can be used for auditing purposes, potentially helping to track data access and identify potential issues.

3.2.4.7 Does the architecture address the data sovereignty?

FIWARE partially addresses data sovereignty, but it relies on other initiatives and implementations to fully achieve it.

- As FIWARE is an open-source platform, it allows organizations to avoid vendor lock-in and have more control over their data and applications.
- Also, since FIWARE promotes interoperability between different components and platforms, this allows organizations to choose and integrate components that align with their data sovereignty needs.

The core Identity Management system (Keyrock) ensures that data access and activities are securely managed and authenticated, which also holds the principles of data sovereignty. Furthermore, the IDS connectors apply data policies and rules in the data when exchanged with different entities to align with the legal requirements.

3.2.4.8 Does the architecture address inter- and intra- data spaces interoperability?

FIWARE Generic Enablers offer interoperability with other systems that adhere to open standards. This ensures data from different sources can be understood and exchanged seamlessly across data spaces.

GE such as the Orion Context Broker, by adhering to open standards like NGSI-LD, allows data exchange with other context brokers in interoperable data spaces [21].

The OLIOT-MG [22] is a specific example of interoperability. It describes how FIWARE can be used as a bridge to connect with systems using GS1 standards, facilitating data exchange between different systems.

FIWARE's modular architecture also contributes to intra-data space interoperability.

3.2.4.9 How well can the architecture scale to handle increasing data volumes and user demands while maintaining efficient performance?

FIWARE's architecture provides several features that enhance scalability [23], enabling it to manage increasing data volumes and user demands while maintaining efficient performance. Thanks to the modularity of components like the Orion Context Broker, these are compatible with and supported by technologies such as Docker or Kubernetes. This support allows for automatic scaling, load balancing, and management of containerized applications, ensuring that the system can scale effectively and handle failures efficiently.

3.2.4.10 Adherence to relevant regulations, e.g., GDPR

FIWARE complies with the current European legislation related to Personal Data Protection [24], Users' Privacy, and the Secrecy and Security of Personal Data, as established in the EU General Data Protection Regulation.

3.2.4.11 Protocols used for communication

The most common protocols used in FIWARE [25]:

HTTP/HTTPS: Widely used for communication between FIWARE components.

NGSI (Next Generation Service Interface): It is a key API standard in FIWARE for context management. It defines a standardized way for components to access and manage context information, regardless of the underlying protocol used for communication.

The FIWARE NGSI API [26] defines:

-a data model for context information, based on a simple information model using the notion context entities.

-a context data interface for exchanging information by means of query, subscription, and update operations.

-a context availability interface for exchanging information on how to obtain context information.

FIWARE NGSI describes three main concepts of the NGSI data models: context entities (entity id, entity type), context attributes (name, type, value), and context metadata (name, type, value).

FIWARE NGSI API specifications have evolved over time, initially matching NGSI-v2 specifications, now aligning with the ETSI NGSI-LD standard.

There are other potential protocols used in FIWARE such as MQTT (for communication between devices and the FIWARE platform in IoT applications), AMQP (for message queuing and reliable messaging between FIWARE components), XMPP (for real-time communication and presence management within FIWARE deployments).

3.2.4.12 Current stage of development

FIWARE is one of the most mature contributors to the data space implementers ecosystem. Components from the FIWARE catalogue [27] have been widely adopted across various sectors and integrated into different projects. Moreover, its modularity by design enables the use of different software pieces independently for a specific service or third-party technology (i.e., AI, blockchain, edge computing, etc.).

Among the most relevant FIWARE components are the NGSI-LD API specifications and the Smart Data Models Initiative.

FIWARE is a mature open-source platform with ongoing development activity. The latest version (FIWARE 8.4.1) [28] was released in December 2023. It includes updates to all NGSI-LD context brokers (Stellio 2.10.2, Orion-LD 1.5.0 and Scorpio 4.1.11). This indicates a commitment to stability and ongoing maintenance of the platform.

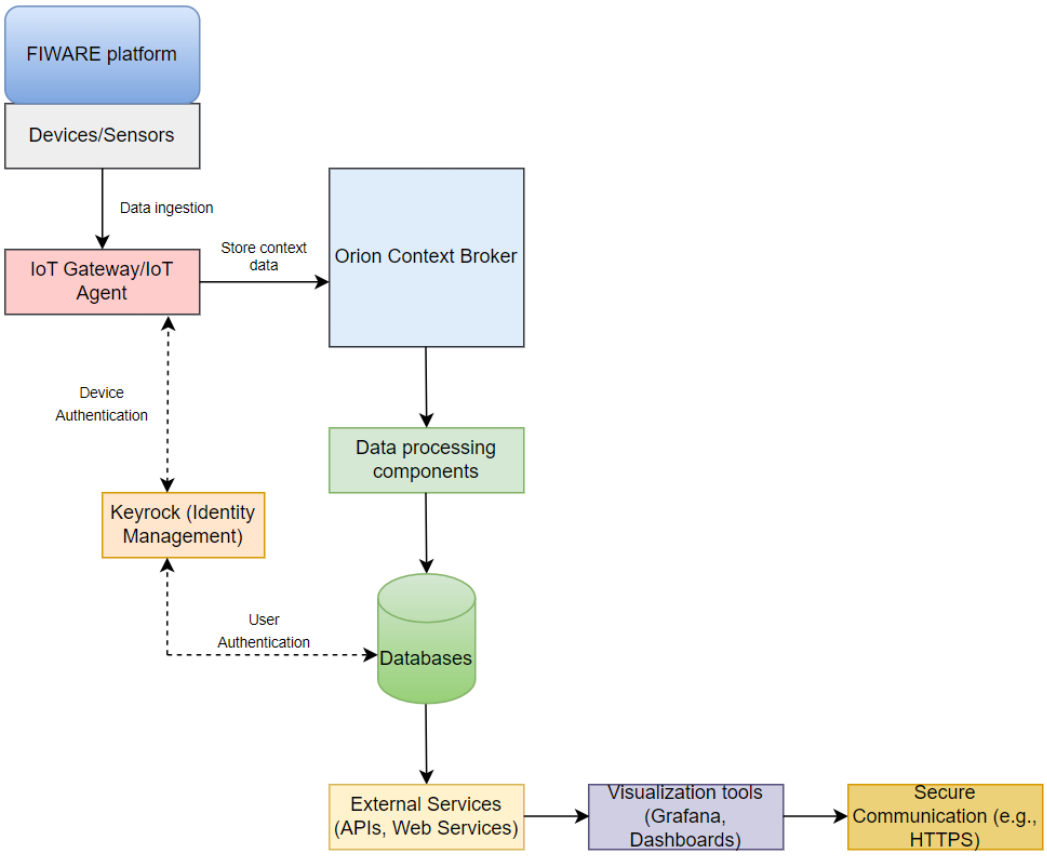

Figure 8. Overview of the FIWARE components landscape.

3.2.5 IHAN

3.2.5.1 Conceptual Foundation

Fair value exchange is at the heart of the whole IHAN ecosystem. Not only must Service Providers be compensated for the creation of the Services but, equally importantly, the Data Providers must be compensated for storing data and making that data available. Value can be money or any other form of value exchange that both sides transparently consider to be fair.

> NOTE: IHAN Blueprint is not something that you take as a complete specification and start developing. Blueprint is a collection of requirements that can be used to design IHAN-compatible components or solutions and an overall description of how the components are arranged and how they interact with each other and the surrounding infrastructure. For instance, we describe **what** the Wallet should do but **not how** it should be done. In another example, we describe what the IHAN Identifier is and what it consists of but not how it should be generated in detail or managed.

Figure 9 Note based on the IHAN latest documentation version[5].

1. **Goals**: The use of digital services constantly generates data that companies use for their benefit, but the use of that data should be fair and transparent. IHAN project built a European data economy model that is aimed at providing a Human-driven European data market, where companies that use data responsibly and open-mindedly succeed with smart services.

The main goals of the IHAN project and its architecture model are:

- Data will be shared more freely between different parties.
- Trust in service providers will encourage individuals to share their data when the sharing is based on their consent.
- People will obtain access to more targeted services that improve their well-being and daily lives.
- Companies of all sizes will achieve growth through innovation and well-being will increase.

The IHAN® blueprint includes the descriptions of the principles and components of IHAN's functional architecture as well as guidelines for building fair data economy services with the aid of existing technology.

2. **Target use cases:** The pilot projects were used to build technical solutions based on the principles of the IHAN operating model. The projects can be found here [28].

Identity management

- Consent management solution (students eligible for an insurance discount)
- Solution for electronic identification
- Component for consent management
- Digital identity standard for music makers
- Mobile wallet for identity management

[5] https://www.sitra.fi/app/uploads/2018/11/261018-ihan-blueprint-2.0.pdf

Health

- Data collection solution for a network (data between pharmacies, patients, controllers and the pharmaceutical industry)
- Using health data abroad
- AI transparency
- Blood glucose measurement and data sharing for children with diabetes
- Using an athlete's personal data in coaching
- Using the well-being data of conscripts to boost physical condition

3. **Data model focus:** In IHAN, the technical data transportation mechanisms are outside of the scope – data and service providers can use whatever technical solutions are fit for their purposes. So, data transfer within IHAN I about different models (direct, aggregator, broker) and requirements.

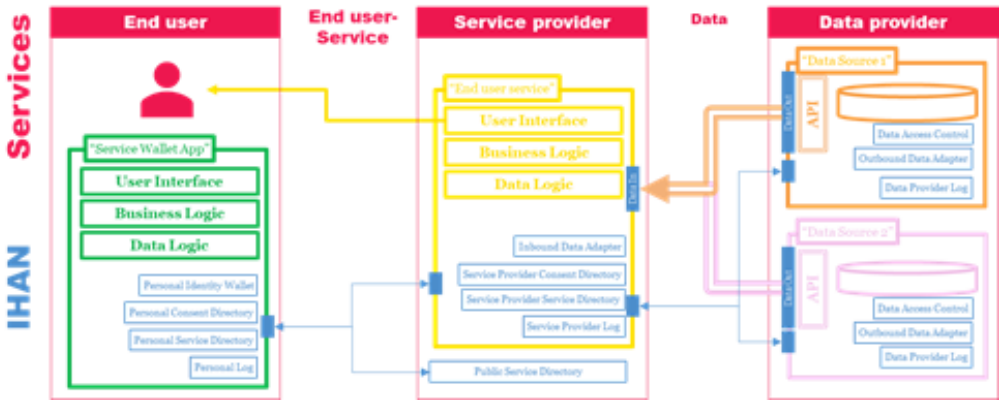

Figure 10 IHAN architecture model

3.2.5.2 Governance mechanisms

While IHAN does not detail a specific set of 'governance mechanisms' by that name, they do describe a governance framework and essential components for building a fair and trusted data ecosystem. These elements, taken together, can be interpreted as mechanisms that seek to regulate interactions within the IHAN data economy. The most relevant of these are described below:

- Consent as the basis for the exchange:

Legal basis (GDPR): IHAN, being based in the European Union environment, is governed by the General Data Protection Regulation (GDPR). Consent, as defined in Article 4 (11) of the GDPR, is crucial for lawful processing of personal data.

Enabling technical component: IHAN requires that consent is not limited to an abstract agreement, but includes the technical means to access the data. This involves providing details of authorisation to access the data from the data provider.

Consent receipt standard: For interoperability, IHAN proposes a common format for consent receipt. This makes it easier for companies to join the ecosystem and for individuals to control the use of their data.

Benefits of traceability: Proper recording of consents increases regulatory compliance and user confidence.

- Roles and responsibilities:

Empowered users: IHAN is user-centric, providing users with tools (such as the Personal Services Portfolio) to manage their identities, consents and access services.

Transparent service providers: These actors need to obtain explicit consent from users to access data and offer their services. They must clearly detail what data they need and for what purposes.

Responsible data providers: They are custodians of data and must ensure that their access and use comply with the GDPR and the consents given. They must facilitate the process of providing data to services in a transparent manner.

- Logging and transparency through logs:

Importance of logs: Logging is a central principle of IHAN to build trust. Every operation related to user data must be immutably and securely recorded.

Proof of respect for the will of the user: Logs provide evidence that users' decisions about their data have been respected.

Different architectural models: IHAN supports centralised, decentralised and distributed models for storing and managing logs.

Log content and access: It details what information should be recorded and who has access to it. A balance is sought between transparency and protection of sensitive data.

- Guiding architectural principles:

Decentralisation: IHAN prioritises decentralised solutions to avoid concentration of power and to promote individual control of data.

User-centric approach: IHAN's architecture seeks to make it easy for users to use and understand, providing intuitive and transparent tools.

Security and compliance: the architecture must ensure data security throughout its lifecycle and compliance with relevant regulations, such as GDPR.

As a resume, governance mechanisms in IHAN are not based on a centralised authority, but on a set of principles, tools and practices that together seek to create a fair, transparent and trustworthy data ecosystem. Consent management, clear role definition, transparent registration and a decentralised architecture are key elements in this model.

3.2.5.3 Security features

While IHAN does not present a specific list entitled ‘security features’, a number of security-oriented mechanisms and principles can be inferred from the description of the IHAN architecture and its components.

- **Consent management as a security pillar:**
  - The informed and explicit consent of the user is the basis for any data access in IHAN.
  - The user is sought to have granular control over what data is shared, with whom and for what purpose.
  - Data portability (a right granted by the GDPR) is facilitated by allowing users to transfer their data between providers
  - A standard format for the receipt of consent is proposed, making it easier for users and providers to manage and understand.
- **Roles and responsibilities for data protection:**

Empowered users: The IHAN architecture provides users with tools to manage their identities, consents and access, giving them greater control over their data.

Service providers: They must be transparent about the use of data and obtain explicit consent from the user. In addition, the architecture limits providers' access to sensitive information, such as user credentials.

Data providers: They are responsible for custodianship of data and ensuring that their access and use is in compliance with the GDPR and consents given. Access control mechanisms are implemented to verify authorisation of requests.

- **Design and architectural principles for enhanced security:**

Decentralisation: Decentralised solutions are promoted to avoid single points of failure and concentration of power that could compromise security.

Focus on privacy by design: The IHAN architecture is designed with data privacy in mind from its conception, seeking to minimise the amount of personal information stored and transmitted.

Regulatory compliance: The architecture must enable the implementation of solutions that comply with data protection regulations, such as GDPR.

Communications security: The use of HTTPS is mentioned to secure communications between decentralised components.

- **Specific security mechanisms:**

The use of encryption to protect sensitive information in consent forms is mentioned. The data provider's part of the form is encrypted in such a way that only the data provider can read it, protecting the user's credentials.

Access to different components, such as identity wallets and service directories, is controlled by authentication mechanisms. Although the sources do not specify authentication levels or methods, it is mentioned that they are required.

Summarizing, the IHAN architecture addresses security through a combination of design principles, access control mechanisms, consent management and immutable logging. It seeks a balance between data protection, transparency and usability, empowering users and promoting a trusted data ecosystem.

3.2.5.4 Standardization

https://github.com/IHAN-Testbed/standardsFirst Published Standard by IHAN community [30]:

> https://sales.sfs.fi/en/index/tuotteet/SFS/CEN/ID5/1/996254.html.stx?_ga=2.259869898.1588871291.1639738429-17692516.1639738429*This document: - defines a digital identity for a natural person for the contextual processing by information systems and machines; - sets the background for all the other components needed to use and utilize the digital identities within a decentralized data economy, such as consent, logging, data transport, services, etc.; - focuses on providing a solid and focused background to deliver a practical approach for future development and still covers the digital identity definitions from a wide enough perspective to not limit its use in today's needs, technologies or industrial use cases; - produces a neutral, objective and generic definition for all humans that can then be scaled up based on the industry, use case and technology it is applied to based on this core definition; - covers also the basic mechanisms for use in the digital services (contextual use), trust and identity management that are within the scope of the digital identity itself; - defines a well-considered overview on the individual's digital properties, their usage and needed core processes for*

further consideration on standardization; and - describes the need for truly decentralized identity for every human being in the digital age.

https://github.com/IHAN-Testbed/standards/blob/master/draft/DataProducts/README.md

- Adoption of existing standards:

RESTful Application Programming Interface (API): IHAN recommends the use of RESTful APIs for communication between different components of the ecosystem. This architectural style, which is widely used in the industry, facilitates interoperability between systems developed by different vendors.

JSON and XML data formats: JSON and XML are mentioned as data exchange formats between components. These formats, which are also widely used, promote interoperability by providing a common structure for information.

HTTPS protocol: IHAN mandates the use of HTTPS to secure communications between the decentralised components of the ecosystem. This industry-standard protocol ensures the confidentiality and integrity of transmitted data.

Digital identity standards (not specified): The need to integrate IHAN with robust electronic identity management systems and digital identity standards, such as SAML, OAuth and OpenID Connect, is mentioned. However, the sources do not specify which of these standards will be adopted or how they will be implemented.

- Creation of common specifications:

Standardised consent receipt: IHAN seeks to define a common format for the consent receipt, which would facilitate interoperability between service providers and allow users to manage their consents more easily.

Metadata to describe services and data sources: The need for service and data providers to register their offerings in the Public Services Directory, including metadata describing the information and accesses required, is described. Although the sources do not specify the format of this metadata, it is inferred that it should follow a common structure to ensure interoperability.

Logs in standard format: IHAN establishes the need to record relevant operations in a standard format, including key information such as the operation performed, the components involved, user information and the time stamp. This standardisation would facilitate auditing, trust building and troubleshooting.

- Promoting interoperability:

Interoperable identity wallets: It is mentioned that there can be several Personal Identity Wallet systems and that they should be interoperable. This would allow users to choose the wallet of their choice without losing access to ecosystem services.

Multiple Public Services Directories with a unified view: Although there may be several physical Public Services Directories, from the user's perspective they are presented as a single logical directory. This facilitates service discovery and interoperability between different directory instances.

There is no IHAN's participation in formal standardisation bodies such as ISO, IEC or W3C.

Overall, it can be said that IHAN recognises the importance of standardisation in building a fair and trusted data ecosystem. However, the information provided in the sources is limited and additional details are needed to fully understand how standardisation will be implemented in practice.

3.2.5.5 Added services such as data anonymization tools, data quality management functionalities

No specific services or functionalities are described apart from the different major components required for decentralized data economy services:

- personal directory
- public directory
- service interfaces
- user preferences

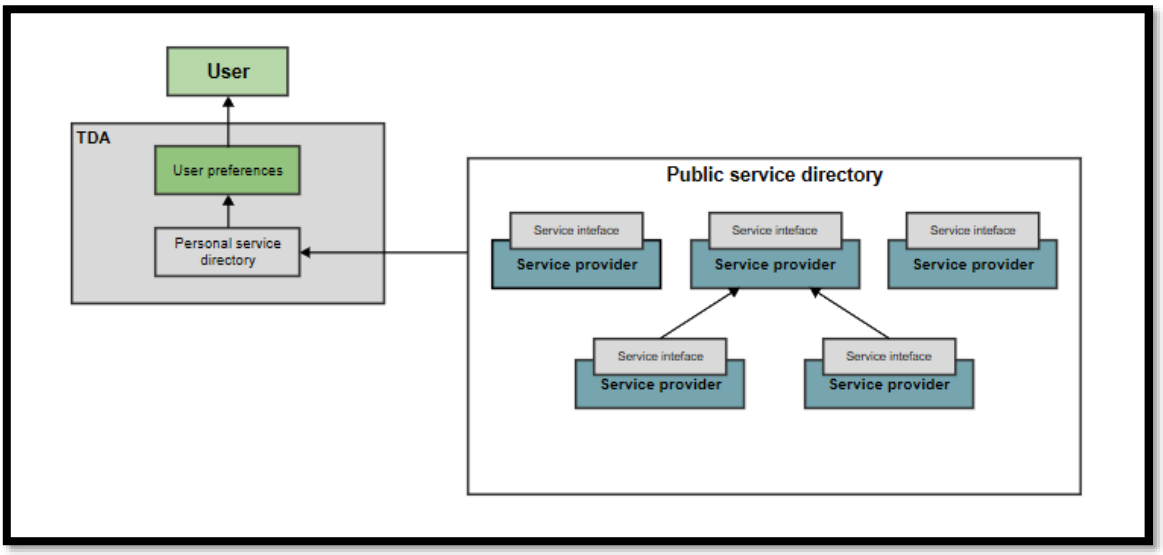

Figure 11. IHAN Service Components diagram

The data which is shared by data sources for the use of services shall be made available using the standard method described by the Data Source components. The standardized format should be aligned with other streams.

The data source offers data based on some terms, which may include SLA's, Consent, Pricing, Logging protocols etc. This should all be standardized in such a way that Service providers can shop for data providers easily, and that it can be abstracted away by service providers, so users don't need to see the complexity.

Any information relevant to the user's consent should be communicated via the user consent channels. The user can act as a data source.

3.2.5.6 Does the architecture offer functionalities for auditing data access and lineage tracking to ensure accountability and compliance with regulations?

The IHAN architecture provides a robust framework for auditing data access and tracking data lineage through immutable, detailed and interconnected logs. This functionality is essential to ensure accountability, compliance and transparency in the IHAN ecosystem.

- Immutable, detailed logs:

  - Key components of IHAN, such as the Personal Identity Wallet, Personal Services Directory, Data Access Control and Inbound and Outbound Data Adapters, generate records of all relevant transactions.
  - These logs include crucial information such as: the operation performed, the component or system that executed it, the user data involved, the timestamp, the status of the operation (success or failure) and the consent used.
  - The architecture emphasises the immutability of logs, meaning that they cannot be modified or deleted once created, ensuring the integrity and auditability of the information.
- Distribution and interconnection of logs:
  - While each component maintains its own logs, the architecture recognises the importance of connecting them in order to obtain a complete view of the data journey.
  - This implies that service and data providers must provide access to relevant logs to users and to each other, either via APIs or via a shared ledger.
- GDPR compliance and privacy:
  - The IHAN architecture prioritises compliance with the General Data Protection Regulation (GDPR) in the management of personal data.
  - Logs are designed to comply with GDPR requirements, avoiding the storage of sensitive information such as identifiers, full personal data or credentials.
- Benefits for auditing and lineage tracing:
  - Accountability: Immutable and detailed logs allow actions on data to be traced back to the responsible entity or user, facilitating auditing and investigation of potential breaches.
  - Compliance: The ability to audit access to data and demonstrate compliance with consents is crucial for complying with data protection regulations such as the GDPR.
  - Transparency: Access to logs by users and interconnection between the different levels of the IHAN ecosystem promote transparency and trust in data management.

3.2.5.7 Does the architecture address the data sovereignty?

The IHAN architecture addresses data sovereignty prominently by placing the user at the centre of the ecosystem and giving them granular control over their data. The key aspects that underpin this assertion are detailed below:

- Control and ownership of data:

Personal Identity Wallet (PIW): This component allows the user to store and manage their own digital identities and Data Access Records, giving them control over what data is shared and with whom.

Personal Services Directory (PSD): The user manages subscriptions to services and can grant or revoke consents for data access at any time.

- Specification and enforcement of data rights:

Consent forms: Users create consent forms to specify what data is shared, for what purpose and under what conditions. The IHAN architecture promotes the standardisation of these forms to facilitate their management.

Data Access Control (DAC): This component on the data provider side verifies the validity and scope of consents before granting access to the data.?

Immutable logs: Immutable logging of operations allows tracking of data access and usage, providing a mechanism for verifying compliance with consents and detecting possible violations.

- Traceability of actions on data:

Distributed logs: The IHAN architecture enables the interconnection of user, service provider and data provider logs, facilitating the traceability of actions on data across the entire ecosystem.

Standardised APIs: The use of standardised APIs for data access and management is promoted, which facilitates the integration of different systems and transparency in operations.

3.2.5.8 Does the architecture address inter- and intra- data spaces interoperability?

The IHAN architecture addresses interoperability in a broad sense by promoting the use of common standards, standardised data formats and unified service management.

3.2.5.9 How well can the architecture scale to handle increasing data volumes and user demands while maintaining efficient performance?

"New technology experiments must ensure performance and scalability"

While the blueprint describes the components and principles of the IHAN architecture, they do not explicitly address scalability or provide details on how the system would handle increased data volumes and user demands while maintaining efficient performance.

3.2.5.10 Adherence to relevant regulations, e.g., GDPR

The IHAN framework emphasizes the ethical use of data and privacy protection through adherence to regulations like the GDPR. The goal is to empower individuals with control over their personal data and facilitate secure data sharing.

3.2.5.11 Protocols used for communication

While the blueprint describes the IHAN architecture and its components, it does not specify the exact communication protocols that would be used between the different elements of the system.

3.2.5.12 Current stage of development

IHAN project was able to implement real-world business pilots using the first data economy testbed.

Some of those pilots are presented below:

Paperless trade finance [31]:

The global trade that moves the goods we consume still relies on paper-based processes that are centuries old. The exporter Wärtsilä and the bank SEB carried out an experiment in trusted digital data exchange in an effort to improve the out-of-date Letter of Credit processes without compromising any of their confidential company data. The results of the experiment were ground-breaking.

https://www.youtube.com/watch?v=ZNwWmU134GA

Seamless customer experience [32]:

As a public funding organisation, Business Finland provides its customers with a multitude of different services and funding instruments. Although relying on the market-leading CRM systems, they have

been unable to provide the experience they would like for their customers. By testing the models and capabilities on the IHAN testbed they were able create a new development path towards interoperable services and seamless customer experience boosted by artificial intelligence.

https://www.youtube.com/watch?v=TfZWRGN8cXU

Digital company service [33]

The Finnish tax administration (Vero) and its partners have long been investigating and contributing to the development process needed to better serve both Finnish and foreign companies in Finland. The data sharing between companies and public and private stakeholders is still far too cumbersome and a lot of resources are wasted by siloed and untrusted data. The IHAN testbed was used to investigate whether the operating environment for companies in Finland could be improved with fully digital and trusted data sharing between the public and private sectors.

https://www.youtube.com/watch?v=E29X3UD8zBw

3.3 Convergence initiatives and projects

3.3.1 DSSC blueprint and building blocks

In recent years, the initiative to co-create and define Data Spaces activities and software services has gained significant momentum, driven by the recognized value and transformative impact these sector-based Data Spaces can have on the data-driven economy. [35]. The launch of the Data Spaces Support Centre (DSSC) aimed at establishing a shared framework of requirements and best practices essential for accelerating the development of sovereign data spaces—a key element in the digital transformation across all industries and research domains.

The DSSC, composed of 12 consortium partners, provides comprehensive support to data space initiatives, particularly ensuring interoperability. The DSSC works towards facilitating the creation of common data spaces that foster data sovereignty, interoperability, and trustworthiness. The DSSC recently launched the Blueprint v1.0 [34], which marks a significant advancement in the development and implementation of data spaces across Europe. This blueprint serves as a comprehensive guide for organizations aiming to establish or enhance data spaces, emphasizing interoperability, data sovereignty, and trustworthiness.

The DSSC Blueprint defines a data space as a distributed system within a governance framework. The blueprint includes a curated set of definitions surrounding the organizational and technical infrastructure for designing and implementing Data Spaces. Additionally, it presents a conceptual model with the Entity-Relationship diagram of a Data Space. Finally, it groups all the functionalities of a Data Spaces in Data Spaces Building Blocks, a set of legal and technical specifications and implementations. The new iterations of the blueprint introduce optimizations for faster implementation of data spaces and ensure they are adaptable to future technological advancements. It also elaborates on the decentralized nature of data space yet allows for a connected data ecosystem and maintaining data sovereignty in a wide range of data-sharing use cases. The blueprint also provides detailed guidance for the design and development of a data space from inception to maturity. This includes the co-creation method with practical instructions for the integration of essential building blocks and a comprehensive toolkit to understand the ecosystem.

The Data Spaces Building Blocks presented in the Blueprint v1.0 are divided between Business or Organizational Building Blocks and Technical Building Blocks. Organizational building blocks relate to the business aspect of data spaces and how their value is created. It includes all the functionalities regarding governance, management, and legal frameworks of a data space. The technical building blocks present the technological aspects of a data space, including software services and processes, to ensure the feasibility of a data space.

Figure 12. Business and organizational building blocks as defined by the DSSC Blueprint v1.0

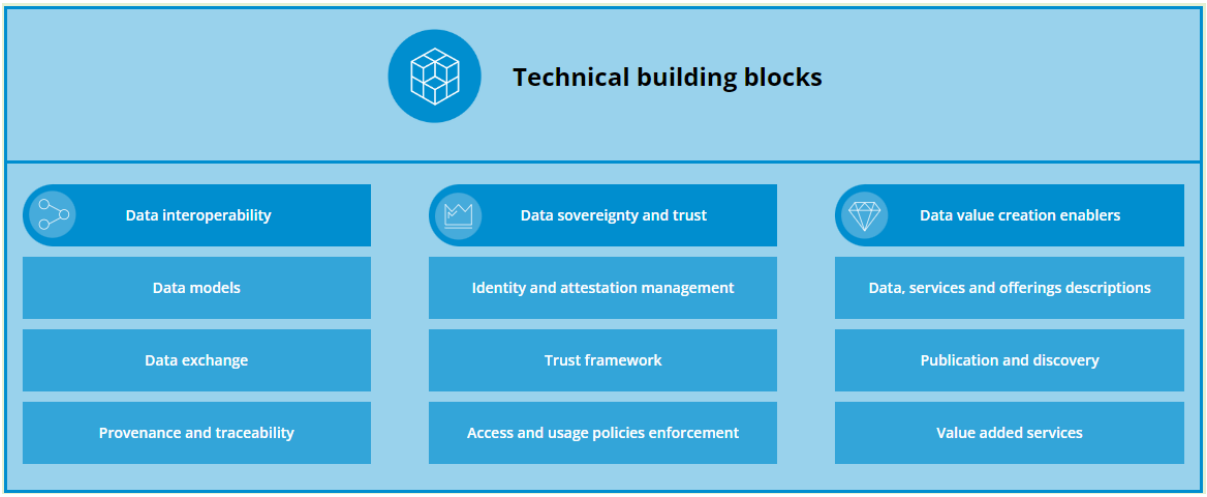

Figure 13. Technical building blocks as defined by the DSSC Blueprint v1.0

3.3.2 DSBA Technical Convergence

The Data Spaces Business Alliance (DSBA) is a collaborative initiative to unite the industry and research players in data-driven actions and data space initiatives. It comprises members from the Gaia-X European Association for Data and Cloud AISBL, the Big Data Value Association (BDVA), the FIWARE Foundation, and the International Data Spaces Association (IDSA). The DSBA published the "Technical Convergence Discussion Document" [36] as a conceptual artifact to define a common reference technology framework for data spaces and as a response to the different isolated specifications and technical pathways that were being generated.

The technical convergence document presents a framework aligned with the existing architectures and models, utilizing shared infrastructure and collaborative implementation efforts. Its primary objective is to ensure interoperability and portability across different data space solutions by harmonizing the necessary technological components. A Minimum Viable Framework (MVF) will be developed to achieve this technical alignment, serving as the foundation for creating data spaces. The initial iteration of the MVF will incorporate a core set of building blocks essential for addressing the three fundamental technology pillars of data spaces: data interoperability, data sovereignty and trust, and data value creation. The document also explores the roles of key components such as the data space connector, data spaces registry, and federated services like marketplaces or metadata brokers.

It explains how these elements can be realized using open industry standards. To illustrate and clarify these concepts, the DSBA provides a detailed example use case with technical descriptions, which can be applied to other scenarios. This use case demonstrates a situation where a data service provider offers a service on a public marketplace, allowing service consumers to access the offering.

The conceptual model present in the published document, introduces the different actors and systems that should take part in a data space ecosystem to achieve the expected functionalities. A Data Space must have a governance authority that governs the data-sharing environment. Participants are members of the data space and are instantiated by participant agents. Every participant is registered in the data space register and verified through a secure and trusted identity provider.

3.4 Table of comparison with standard data spaces architectures

Table 2. Comparison of the revised data space architectures and initiatives.

Feature	**IDS-RAM**	**GAIA-X**	**FIWARE**	**IHAN**
Technical Architecture	Open architecture, reliable and federated for cross-sector data exchange	Federated and open data infrastructure	Standardized APIs	Distributed and open standard-based
Focus	Trusted and sovereign data exchange across sectors	Domain-agnostic (Uses federations for domain-based deployments)	Domain-agnostic	Fair data economy
Standardization	Open-standards-based and focused on promoting interoperability	Compliance framework (data sovereignty, interoperability)	Open standards-based, promotes interoperability	Open standards, promotes interoperability
Deployment Model	Supports a hybrid deployment model, adaptable to participants' needs	Federated	Flexible	Hybrid model, allowing flexibility
Security	IAM, encryption, IDS-certified trusted connectors	Trust Framework (IAM, Multi-factor), GAIA-X register, Built-in eIDAS Cybersecurity measures	IAM, eIDAS Trust Anchor, iSHARE Trust Anchor	GDPR-compliant consent management
Governance Model	IDS Policy Enforcement, ODRL Profiles, eDC Policy Engine	Federated, Compliance service, GAIA-X Registry	Decentralized, governed by the community	User-centric governance model
Scalability	Highly scalable, with a modular architecture that accommodates various organizations	Achieved through the distributed architecture. Available for cloud/edge deployments and containerization services.	Easy integration of additional components and services through modular architecture and open API	Enables both small and large-scale implementations while maintaining user control and data sovereignty

4 Data Sharing Initiatives and Organizations

In addition to the European data space architectures explored previously, several initiatives and organizations have emerged to facilitate secure and trustworthy data sharing across borders. Unlike data space architectures that define technical specifications, these initiatives focus on fostering collaboration and establishing best practices for data exchange. Each initiative offers a unique approach to data governance, contributing to a robust and collaborative European data space [37].

4.1 Data Sharing Coalition

The Data Sharing Coalition [38] is an international initiative in which a wide variety of organizations collaborate to allow data sharing between existing data spaces. By enabling interoperability between existing and future data spaces with data sovereignty as a core principle, parties from different sectors and domains can easily share data with each other, unlocking significant economic and societal value. The Data Sharing Coalition aims to build on existing data sharing initiatives to strengthen them in unlocking the value of sharing data in and across their domain. It also aims to stimulate cross-domain data sharing under the control of the rightsholder, by enabling interoperability between domains.

The Data Sharing Coalition has two clear focus areas: the development of the Data Sharing Canvas as a kind of reference architecture for sharing data and supporting it with the insights and experiences from use cases.

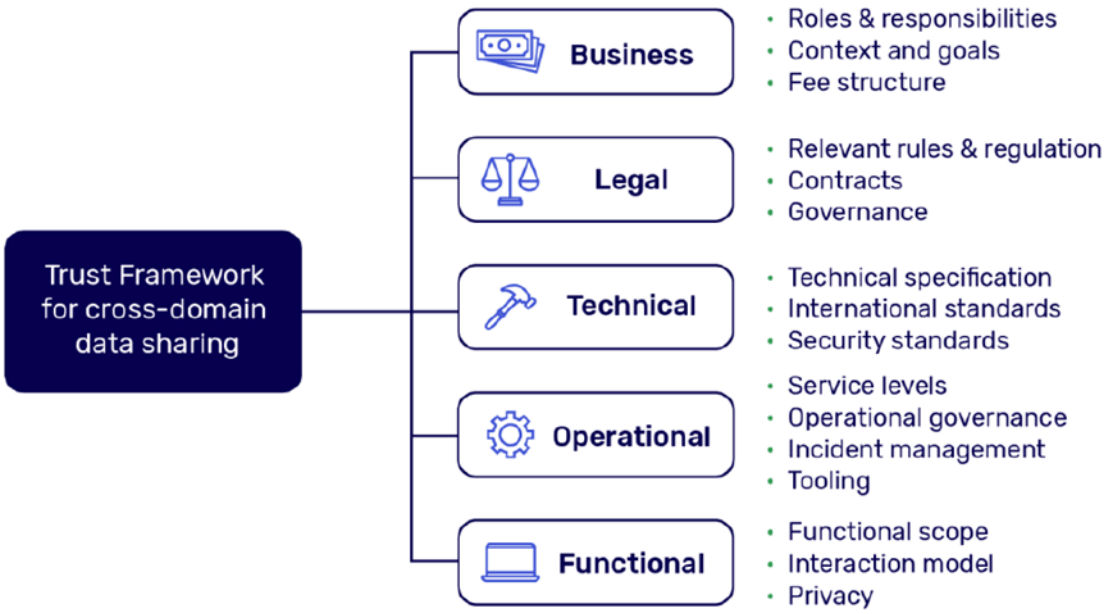

Figure 14 BLOFT model for DSC trust framework[6]

4.2 MyData

MyData is an international non-profit organization founded in 2018. Its purpose is to enhance individuals' rights regarding the use of their personal data. The organization aims to align digital

[6] https://coe-dsc.nl/wp-content/uploads/2024/02/data-sharing-canvas.pdf

human rights with personal data protection, establishing trust between people and organizations. It promotes a user-centric model designed to serve the needs of individuals.

What's MyData:

- A way of thinking: Human-centric, ethical data treatment.
- Interoperable: Hosted by several parties or even self-hosted.
- Constantly evolving concept: Aspiring a global accessible networking.

What is not MyData:

- A single project: Many projects with different focus.
- Geographically limited: A lot of hubs all over the world.

Certain entities need to strengthen the technical and ethical infrastructure to ensure that the MyData principles are upheld. These entities are known as MyData Operators.

The operators follow the MyData Declaration architecture to ensure that the personal data is safe.

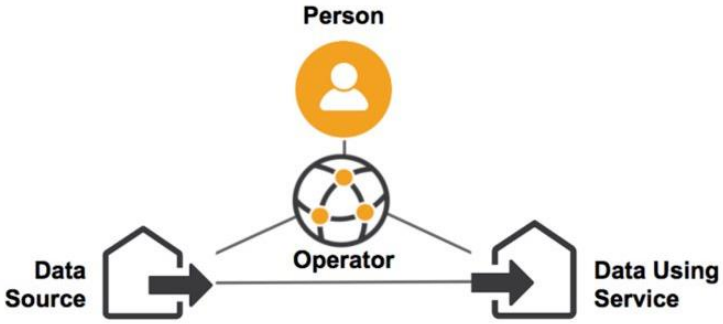

Figure 15 MyData Declaration overview.

The MyData Declaration focuses on practical ways to make personal data management easier and more user-centric. The goal is to enable different operators to work seamlessly together.

A MyData operator must follow the following 4 key points:

1. MyData Operator Blueprint: Defines what MyData operators do and what to expect from them.
2. Making MyData Work Together: Ensures different MyData operators are compatible.
3. MyData Ecosystem Rules: Discusses legal and voluntary frameworks for data governance.
4. MyData Business Models: Explores sustainable business models for user-centric data management.

MyData is not a Data Space in itself, but rather a complement that can bring an individual-centric approach and a robust framework for personal data management within a Data Space

Together, MyData and Data Spaces can coexist and reinforce each other to create a more equitable, transparent, and individual-centric data ecosystem.

4.3 Big Data Value Association (BDVA)

BDVA [39] is an industry-driven research and innovation organization with a goal to develop an environment that enables the data-driven and AI-enabled advancements of the economy and society in Europe.

The organization has the following objectives:

- Create a positive impact on policy-making, business and society through the strategic use of data and AI technologies.
- Keep up with the dynamic change that AI and Data bring to business and society.
- Ignite Data and AI world-class research to enhance European competitiveness on the global stage.
- Foster collaborative innovation by co-creating new projects and maximizing their potential impact.
- Contribute to a sustainable future.

BDVA facilitates the collaboration of existing regional partners at the European level through the provision of resources and expertise to promote the co-creation, development, and experimentation of pan-European data-driven applications and services [40].

BDVA was established in 2014 as the private counterpart of the European Commission in the Big Data Value Public Private Partnership (BDV cPPP) to support the development of the European Big Data Value ecosystem and to provide a unified voice for the European big data community.

The organization has also established a robust and expanding alliance with GAIA-X, IDSA and FIWARE through the Data Spaces Business Alliance (DSBA) and some national initiatives for Industrial AI.

4.4 The Data Competence Center for Cities and Regions (DKSR)

The Data Competence Center for Cities and Regions [41] (DKSR) supports municipalities and municipal enterprises in the use of data-based applications for urban and regional development. They achieve this by focusing on three key areas:

- Building data competence: DKSR offers consulting services to help cities develop the skill and knowledge to leverage data for sustainable development, efficient management, and achieving specific city goals.
- Urban Data Community: The Urban Data Community connects cities (municipal stakeholders) to share best practices and collaborate on projects. This allows successful data-driven solutions to be implemented in multiple cities and adopted more widely.
- Open Urban Data Platform: The DKSR OUP is an open-source urban data platform that acts as a central hub for collecting, integrating, and analyzing data from various sources within a city. The platform has its particular strengths operating IoT-based applications with real-time functionalities, which allows for a more comprehensive understanding of how a city functions. Its core strengths lie in:
    - Open source: Fosters transparency and collaboration. This means that everyone can freely access the platform's source code, ensuring that cities and regions do not become dependent on one vendor.
    - Data sovereignty: The OUP prioritizes data ownership by cities. Data is monitored and controlled via the integrated IDS Connector as a user interface.
    - Standardized: The Open Urban Platform (OUP) of the DKSR is a (near-) real-time data platform that follows the vision of open urban platforms as expressed by the European Innovation Partnership Smart Cities and Communities (EIP-SCC) and defined as a standard in the DIN SPEC 91357 framework.

- Scalability: The platform is designed to be scalable, accommodating the needs of both small towns and large metropolises. This eliminates the need for significant infrastructure changes as a city grows.

  Also, the DKSR OUP is FIWARE certified.

- Real-time capability: The DKSR OUP can handle real-time data streams, enabling cities to monitor and react to events as they happen. This real-time data is crucial for applications like traffic management, emergency response, and environmental monitoring.

- Data integration: The platform offers a connector technology that can link heterogeneous data sources from all municipal sectors and then homogenize them. This allows for more informed decision-making by city officials.

- https://www.dksr.city/en/the-dataplatform/.https://www.dksr.city/en/the-dataplatform/

5 Drawbacks and advantages

Table 3. Summary table with main advantages and drawbacks of each data space architectures evaluated.

		Advantages	Drawbacks
IDS-RAM	**Technical**	Data sovereignty Interoperability Security mechanisms Standardization	Complex implementation of the architecture Limited availability of ready-to-use tools and connectors, requiring custom development
	Legal	Intellectual property protection Contractual framework	Legal complexity increases with cross-border data sharing and compliance requirements
	Economic	Reduces costs by enabling secure and efficient data sharing without centralized intermediaries Market opportunities	Ongoing maintenance costs
	Societal	Industrial efficiency Innovation Sustainability	Dependence on technology
GAIA-X	**Technical**	Data sovereignty Interoperability Federated infrastructure	Dependence on standards
	Legal	Data privacy Trustworthiness	Dependency on EU members common standarization Global scale aimed approach
	Economic	Potential data-driven business models	Complex governance structure Dependency on SMEs implementation
	Societal	Innovation	Ethical concerns related to data usage

		Digital sovereignty	
FIWARE	**Technical**	Open source Interoperability Modular architecture Flexibility	Complex to set up and manage for non-technical users.
	Legal	Supports compliance with GDPR through data management features and consent mechanisms. Promotes data sovereignty and control over personal data through federated architecture.	Open-Source management of IP rights
	Economic	Cost-effective Scalability	Ongoing maintenance
	Societal	Community-driven development Interoperability	Digital divide Data privacy
IHAN	**Technical**	Healthcare-specific architecture User-centric	Complexity of implementation Dependency
	Legal	User rights Promotes transparency and accountability in data usage	
	Economic		Potentially high implementation and compliance costs for business transitioning to IHAN standards Uncertain economic benefits due to limited adoption and recognition of the framework
	Societal	Improved patient care Healthcare efficiency	
	Technical	Standardization	Technical overhead

Data Sharing Coalition		Interoperability	Security risks
	Legal	Compliance	Liability issues
	Economic	Efficiency	
	Societal	Collaboration Transparency	Public trust
MyData	**Technical**	Privacy by design User control Interoperability	Scalability challenges
	Legal	User empowerment Compliance	
	Economic		
	Societal	Transparency	Adoption rate
BDVA	**Technical**	Innovation Standards and Frameworks	
	Legal	Guidelines Regulatory Alignment	Regulatory lag
	Economic		
	Societal	Public services	Privacy concerns Data misuse
DKSR	**Technical**	Scalability Interoperability	Data standardization Technical complexity
	Legal	Data governance	Liability issues
	Economic	Cost efficiency Economic development	
	Societal	Public participation	Privacy concerns

6 Security and privacy gaps

6.1 Security gaps

Data spaces, designed to facilitate secure and trusted data sharing across multiple organizations, face significant security challenges due to their decentralized nature and emphasis on collaboration. While techniques like Secure Multiparty Computation (SMPC) offer privacy-preserving computation, their integration into data spaces is complex and not widely adopted yet [42]. Ensuring interoperability across heterogeneous systems remains a key issue , while the orchestration of IoT, edge, and cloud resources is prone to vulnerabilities[43] [44], while the orchestration of IoT, edge, and cloud resources is prone to vulnerabilities [42]. Securing distributed infrastructures such as edge devices and cloud platforms imposes high costs, particularly as the rapid evolution of AI technologies necessitates scaling security solutions [45]. Furthermore, the increasing adoption of cloud-native technologies introduces new risks, such as data leakage due to misconfigurations or unauthorized access to cloud storage.

This subsection explores some of the key security vulnerabilities that organizations must address to ensure the integrity and confidentiality of data within these environments

6.1.1 Technical challenges

- Identity and Access Management

In the context of data space architectures, effectively managing identities and access across various entities and domains poses a significant security challenge. This is particularly complex in decentralized environments, where multiple stakeholders are involved and have their own identity management systems and security standards. The decentralized nature of identity management can result in inconsistencies in access control policies, which increases the risk of data breaches and illegal access. Furthermore, it can be difficult and time consuming to scale the implementation of complex models such as attribute-based access control (ABAC) to address diverse access requirements. Therefore, in such environments, ensuring that sensitive data is only accessed by authorized entities becomes a challenging endeavour.

- Data Integrity and Confidentiality

Insufficient encryption practices or mishandled key management can leave data vulnerable to breaches, unauthorized access, or tampering. Safeguarding data throughout its entire lifecycle poses a challenge in environments where data may pass through multiple points and undergo numerous transfers and processing events. Effective encryption strategies, robust key management, and consistent monitoring play a vital role in addressing these risks.

6.1.2 Organizational challenges

- Data Sovereignty and Regulatory Compliance

Data sovereignty and regulatory compliance are critical issues given the global nature of many data spaces. Different regions have their own data protection laws, for instance, the GDPR in Europe and the CCPA in the United States, that impose strict conditions for the handling, storage, and transfer of data. These diverse legal frameworks can pose challenges to secure data exchange, especially when there are conflicting regulations related to data localization and usage rights. Maintaining compliance with multiple, and at times contradictory, regulations require careful management of data usage rights, deletion protocols, and data portability, all of which can be technically demanding and resource intensive.

- Governance and Legal challenges

Involving numerous entities from different jurisdictions, each with distinct roles and responsibilities, can create ambiguities around liability, accountability, and data ownership. In cases of data breaches or misuse, determining who is responsible can be difficult, especially in the absence of standardized governance frameworks. Establishing clear governance structures, clearly defining roles and responsibilities, and developing consistent procedures for handling security incidents are critical for managing these challenges effectively.

6.1.3 Economic challenges

- Interoperability and Standardization

The use of diverse systems and protocols, without a standardized approach to security, can create weaknesses in the data exchange process. Security gaps may emerge from inconsistencies in the security capabilities or configurations of different systems. A lack of unified security standards across all participants complicates the protection of data, underscoring the need for common security frameworks and protocols to minimize vulnerabilities.

- Monitorization and Threat Detection

Given the dynamic and distributed nature of data space architectures, continuous monitoring is essential to maintaining security. Real-time monitoring helps detect unusual activities, prevent data breaches, and ensure that data sharing complies with regulatory requirements.

The decentralized structure of these environments can make it difficult to maintain a comprehensive view of all activities, especially when multiple entities and systems are involved. Ensuring that all data exchanges, transfers, and access points are being monitored without creating blind spots is a complex task. Additionally, processing and analyzing large volumes of log data in real-time can strain resources and require sophisticated tools and techniques.

6.2 Privacy gaps

In the context of data spaces, privacy refers to the protection of personal and sensitive data when it is shared, stored, or processed across different participants of the ecosystem. As data spaces aim to facilitate seamless data exchange and collaboration, they must also ensure that privacy principles such as data sovereignty, confidentiality, and user control are maintained. In recent years, Europe has also presented extensive work on regulations like the General Data Protection Regulation (GDPR) or the Data Governance Act to set high standards for data privacy and protection.

Approaches to ensuring data privacy within a data-sharing ecosystem usually encompass both technical solutions and governance measures. In the technical part, several investigated technologies in this report and initiatives already contain modules for anonymization, encryption and multifactor access. Data space connectors use secure connections and traceability methods to maintain data privacy during data transfers. Data spaces infrastructures have built-in security features such as identity management and encryption mechanisms to maintain data protection and privacy while promoting interoperability. In some cases, standardized APIs and data exchange protocols allow organizations to implement their own privacy controls. Data privacy is involved in the data transfer and the mechanisms for traceability of the data space transactions. Additionally, data space rulebooks and guidelines usually include governance frameworks, consent management, and data sovereignty protocols. Governance frameworks define data-sharing rules, access protocols and compliance with legal requirements and privacy norms. Several data space initiatives use auditing and lineage tracking functionalities to ensure compliance with privacy regulations and trust among the participants of the data-sharing ecosystem. Privacy-preserving data spaces have made considerable advancements in the last few years. However, new challenges and gaps continue to emerge due to the evolving nature of the technology, regulatory changes, and increased data-sharing demands.

Despite the progress in ensuring privacy in data spaces, several challenges still need to be addressed to adapt to the evolving and growing landscape.

- **AI Governance and Explainability:** There is a need for more explainable AI models that can demonstrate how decisions are made, especially when processing personal data and ensuring compliance with GDPR's "Right to Explanation"." [46]. This requires integrating privacy-preserving techniques into data spaces to maintain data confidentiality while still enabling AI-driven insights [47]. Some approaches currently being investigated include but are not limited to Federated Learning or Differential Privacy models.
- **Localization and Data Sovereignty:** Decentralized data storage and processing will be essential to ensure data sovereignty is respected across data spaces **Localization and Data Sovereignty:** Decentralized data storage and processing will be essential to ensure data sovereignty is respected across data spaces [48]. As more countries enforce sovereignty and localization laws, data spaces face challenges in ensuring that data remains under the control of its owner and is stored and processed within specific jurisdictions.
- **Regulatory Compliance and Cybersecurity:** Data spaces must implement robust and safe cybersecurity measures, regular risk assessments, and comprehensive privacy governance to remain compliant with evolving regulatory frameworks. They should also adopt a security-by-design approach and improve the detection of anomalies and threads. Due to the rapid evolution of regulatory actions and laws, the strict enforcement of existing laws and new directives will be a major challenge for data space initiatives. [49].
- **Unstructured data sharing:** The emergence of methods and processing techniques that can get insights from unstructured data such as video or images by AI models creates significant privacy risks, as they may contain sensitive information that is difficult to manage and protect **Unstructured data sharing:** The emergence of methods and processing techniques that can get insights from unstructured data such as video or images by AI models creates significant privacy risks, as they may contain sensitive information that is difficult to manage and protect [50]. There is a requirement for advanced tools to classify, secure and anonymize unstructured data within data spaces, ensuring that sensitive data is detected and remains protected.
- **Data protection and online safety for groups at risk:** Protecting children's, minorities, and other risk groups' data online has become a European priority, especially given concerns about the misuse of personal data on social media. [51],[54]. Data spaces need to incorporate stricter verification measures and privacy controls to protect minors and ensure compliance with design standards and regulations.

Privacy and regulatory compliance in data spaces is a dynamic and evolving field that requires both technical and organizational approaches. While current implementations already target the challenge of keeping a privacy-preserving, interoperable ecosystem, challenges such as sovereignty of data, AI governance, unstructured data, and online safety present new gaps that must be addressed in the next steps of the data spaces initiatives. Privacy-preserving implementations will improve transparency and ethical data practices and evolve data spaces into secure and trusted collaborative environments.

7 Data spaces interoperability gaps

Interoperability in data spaces is crucial for the effective functioning of data-driven ecosystems, especially in Europe, where diverse organizations and systems must exchange and utilize data seamlessly [52][53]. These data spaces are designed to enable secure, sovereign, and scalable data sharing among businesses, governments, and individuals. However, achieving interoperability is challenging due to the wide range of technical, semantic, organizational, and legal differences across various domains and regions.

To address these challenges, significant standardization efforts are underway. The European Interoperability Framework [55] (EIF) provides a set of recommendations and guidelines to enhance interoperability among public administrations and other organizations within the EU. The EIF defines interoperability at four levels (see Figure 16): legal, organizational, semantic, and technical, offering a comprehensive approach to overcoming interoperability barriers.

Moreover, standards like ISO/IEC 19941 [56] on Cloud Computing Interoperability and Portability aim to provide technical specifications to facilitate interoperability in cloud environments (see Figure 17).

The initiatives described in Chapter 3, such as IDS-RAM, FIWARE, GAIA-X, and IHAN, contribute to achieving interoperability by adopting and promoting these frameworks and standards. For instance, IDS-RAM incorporates the Dataspace Protocol (DSP), which is on its way to becoming an ISO standard, highlighting its importance and impact on interoperability. These architectures align with the EIF's recommendations and work towards common interoperability solutions, as detailed in the comparative analysis in 3.4.

Despite these efforts, the process of fully realizing interoperability across all facets is still ongoing. Continuous collaboration, standardization, and convergence initiatives are essential to overcome existing gaps and achieve seamless data sharing across diverse systems and organizations.

- Technical: Difficulty in aligning diverse technical standards and ensuring systems can connect and exchange data effectively.
- Semantic: Challenges in achieving consistent understanding and interpretation of data across different domains.
- Organizational: Complexities in coordinating processes, policies, and governance structures between various organizations.
- Legal and Regulatory: Issues in harmonizing legal frameworks and ensuring compliance with regulations across different jurisdictions.

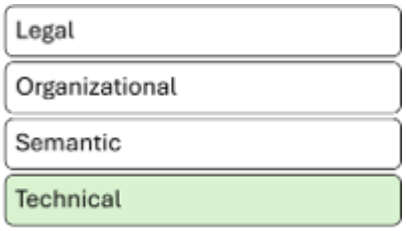

Figure 16. European Interoperability Framework

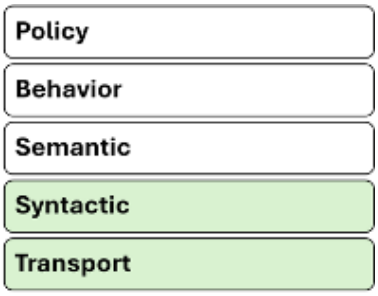

Figure 17. ISO19941 - Cloud Computing Interoperability and Portability.

7.1 Technical Interoperability Challenges

Technical interoperability forms the core of data spaces, enabling diverse systems, platforms, and devices to communicate and exchange data effectively. Achieving technical interoperability is essential for promoting and exploiting collaboration across sectors, ensuring seamless data sharing regardless of the underlying technologies. However, based on the Data Spaces analysed in this document, several challenges must be highlighted on the path to establishing robust technical interoperability.

To facilitate the identification and management of potential Technical Interoperability Challenges, we have categorized them into specific domains [57]:

- Protocol Standardization Issues (TI1): Different data space architectures often use varying communication protocols. Aligning these protocols is key for smooth system interactions.
- Data Format Compatibility Challenges (TI2): Data comes in many formats across different architectures. Making sure these formats play well together is crucial for effective data sharing.
- Identity and Access Management Integration Difficulties (TI3): Each architecture might handle user authentication and authorization differently. Finding common ground here is essential for secure cross-architecture operations.
- Connector Technology Interoperability Problems (TI4): Various architectures employ different connector technologies for data exchange. Ensuring these connectors can work together is vital for a unified data ecosystem.

Table 4. Technical interoperability challenges along considered Data Space Architectures.

		IDS-RAM	FIWARE	GAIA-X	IHAN
Main treat		Data Space Protocol (DSP) and Trusted Connectors	NGSI-LD API and Context Broker	Flexible protocols and Federated Catalogs	Flexible data transportation and connectors
	Context	Emphasizes secure, standardized data exchange. Trusted Connectors for data sovereignty. DSP is being standardized as ISO standard. [59][59] Information layer of IDS-RAM provides a common information model and vocabulary.	Utilizes NGSI-LD API for context information management. Employs Context Broker for real-time data integration	Allows diverse protocols based on industry needs. Facilitates service and data discovery across the ecosystem	Supports various protocols and connectors, focusing on user data control and ethical data management
	Potential Gap	TI3, TI4 - The DSP addresses protocol standardization (TI1) and data format compatibility (TI2). However, integration challenges in Identity and Access Management (IAM) (TI3) may still exist due to complexities across diverse organizations. Additionally, connector interoperability (TI4) remains a concern, as variations in connector implementations by different vendors could lead to issues, despite DSP providing behaviour specifications.	TI1, TI2, TI4 - FIWARE-specific components may face integration challenges with non-FIWARE systems	TI1, TI2, TI3, TI4 - Flexibility could lead to fragmentation and inconsistencies across different domains	TI1, TI2, TI3, TI4 - Flexibility may reduce predictability and standardization, complicating broad interoperability

7.2 Semantic Interoperability Issues

As data space architectures evolve, semantic interoperability becomes increasingly critical. This involves overcoming challenges that hinder effective data exchange across platforms. Important Semantic Interoperability (SI) Challenges have already been pointed out [60], as well as potential solutions to overcome these [61]. Important challenges are:

- Vocabulary and Ontology Alignment (SI1): Each architecture may use different terminologies or ontologies, making it challenging to align vocabularies across data spaces.
- Context Preservation (SI2): Ensuring that the context and meaning of data are preserved when shared across different data spaces and domains remains a significant challenge.

- Cross-Domain Semantics (SI3): While general semantic models exist, integrating domain-specific semantics , mobility [58]consistently across different data space architectures is complex [62].
- Metadata Standardization (SI4): Although architectures like IDS-RAM emphasize metadata semantics, ensuring consistent interpretation and use of metadata across different data it needs to be considered within the rest of the Data Space Architectures.
- Evolution of Semantic Models (SI5): As data spaces evolve, keeping semantic models up-to-date and ensuring backward compatibility becomes a maintenance task to be considered.
- Cross-Domain Interoperability (SI6): Like in SI3, achieving semantic interoperability across different domains remains a significant challenge, as domains may have its own vocabulary.

Specific Semantic interoperability issues given the studied Data Space architectures can also be seen in Table 5.

Table 5. Semantic interoperability issues along considered Data Space Architectures.

		IDS-RAM	FIWARE	GAIA-X	IHAN
Main treat		Common Information Model and Vocabulary Provider	NGSI-LD API and Smart Data Models	Flexible ontology and metadata standardization	User-centric and fair data economy focus
	Context	Common Information Model [62] for shared concepts & data structures. DCAT for metadata cataloguing, ODRL for data rights. Vocabulary Provider [64] for domain-specific vocabularies, aligning . Aligns with the EIF's semantic interoperability.	NGSI-LD API is based on linked data principles and promotes standardized data models through the Smart Data Models initiative [63]	Aims to promote ontologies for interoperability within and across sector-specific data spaces. Focuses on metadata standardization to accommodate diverse industry needs	Adopts a flexible approach to data transportation mechanisms. Emphasizes individual control over data and fair data economy principles
	Potential Gap	SI1, SI2, SI3, SI5 - Challenges in aligning with external vocabularies and cross-domain integration	SI1, SI3, SI4, SI5 - Potential misalignment with non-NGSI-LD systems and domain-specific standards	SI1, SI2, SI3, SI4, SI6 - Flexibility may lead to inconsistencies across sectors and complicate model evolution	SI1, SI2, SI3, SI4, SI6 - User-centric approach may create challenges in standardization and cross-system interpretation

7.3 Organizational Interoperability Barriers

Organizational interoperability in data space architectures presents complex challenges beyond technical considerations. The European Union's General Data Protection Regulation (Art. 32, GDPR) [65] emphasizes the importance of robust organizational measures for data protection and security, highlighting the need for careful consideration of compliance and data sharing agreements (specially in terms of Participant Onboarding and Compliance and Data Sharing Agreements). As IoT ecosystems increasingly intersect with data spaces, frameworks addressing both semantic and organizational interoperability become crucial [66], considering governance model alignment, standardization of onboarding processes, and harmonization of operational workflows (introducing Governance Model Alignment and Operational Process Harmonization). These challenges are further explored in comprehensive studies on designing data spaces [67] which emphasize the ecosystem approach to competitive advantage. The following points outline four key Organizational Interoperability (OI) challenges in data spaces:

- Governance Model Alignment (OI1): Challenges in aligning different governance structures and decision-making processes across data spaces.
- Participant Onboarding and Compliance (OI2): Difficulties in standardizing the process of integrating new participants and ensuring their compliance with data space rules.
- Data Sharing Agreements (OI3): Complexities in establishing and enforcing consistent data sharing agreements across different organizational cultures.
- Operational Process Harmonization (OI4): Challenges in aligning operational processes and workflows across different data space participants.

Table 6. Organizational Interoperability issues along considered Data Space Architectures.

		IDS-RAM	FIWARE	GAIA-X	IHAN
	Main treat	Decentralized governance model	Open-source community governance	Federated services and self-description	Fair data economy principles
	Context	Emphasizes participant autonomy within a common framework. Uses Clearing House for transaction logging and clearing	Relies on a foundation-led governance model with community input. Focuses on open standards and open-source implementations	Implements a federated governance approach with strong emphasis on European values and data sovereignty	Promotes user-centric data management and ethical data use through a collaborative governance model
	Potential Gap	OI1, OI2, OI4 - Decentralized model may complicate decision-making and process alignment across different IDS deployments	OI2, OI3 - Open nature may lead to challenges in enforcing strict compliance and standardized data sharing agreements	OI1, OI2, OI3 - Federated approach might result in inconsistent governance practices across different domains and regions	OI2, OI3, OI4 - User-centric focus may create challenges in aligning with traditional organizational processes and agreements

7.4 Legal and Regulatory Interoperability Constraints

Legal and regulatory interoperability presents significant challenges in the development and implementation of data space architectures. As data spaces evolve to facilitate seamless data exchange across diverse organizations and jurisdictions, they must navigate a complex landscape of legal frameworks and regulatory requirements. Previous work done by [68] highlight the intricate challenges of GDPR compliance in health information exchanges, emphasizing the need for robust data protection measures in interoperable systems. Other efforts in the subject suggest user-centric network model for data control [69], addressing the critical aspects of data sovereignty, trust, and security in interorganizational data sharing. Furthermore, [70] underscore the tension between data sharing imperatives and data protection legislation in smart city development, illustrating the broader societal implications of legal and regulatory interoperability. Against this backdrop, data space architects and participants must address several key challenges to ensure compliance, trust, and enable effective data sharing within legal boundaries:

- Legal Framework Alignment (LR1): Challenges in aligning with diverse legal frameworks across jurisdictions and sectors.
- Data Protection Compliance (LR2): Issues related to compliance with data protection regulations like GDPR.

- Data Sharing Agreements (LR3): Complexities in establishing legally binding and interoperable data sharing agreements.
- Regulatory Compliance (LR4): Challenges in meeting various sector-specific regulatory requirements.

Table 7. Legal interoperability issues along considered Data Space Architectures.

		IDS-RAM	FIWARE	GAIA-X	IHAN
	Main treat	Usage control and data sovereignty	Open source and standardization	European values and data sovereignty	Fair data economy and user control
	Context	Emphasizes legal compliance and data sovereignty through its architecture and governance model. Uses Usage Control to enforce legal and contractual requirements	Focuses on open standards and open-source implementations, which may simplify some legal aspects but also create challenges in regulated sectors	Implements a federated approach with strong emphasis on European values, GDPR compliance, and data sovereignty	Promotes user-centric data management and ethical data use, with a focus on fair data economy principles
	Potential Gap	LR1, LR3 - Decentralized model may complicate legal alignment across different jurisdictions. Usage control implementation may face challenges in diverse legal contexts	LR2, LR4 - Open nature may lead to challenges in ensuring consistent regulatory compliance, especially in highly regulated sectors	LR1, LR2, LR4 - Federated approach across Europe may face challenges in harmonizing diverse national regulations and sector-specific requirements	LR2, LR3 - User-centric focus may create complexities in establishing standardized, legally binding data sharing agreements across different jurisdictions

8 Identification of advancements on the existing frameworks

8.1 IDS-RAM

RAM 5 will be aligned with the latest developments in IDSA and Data Spaces. The vision for IDSA RAM5.0 is identified in the picture below:

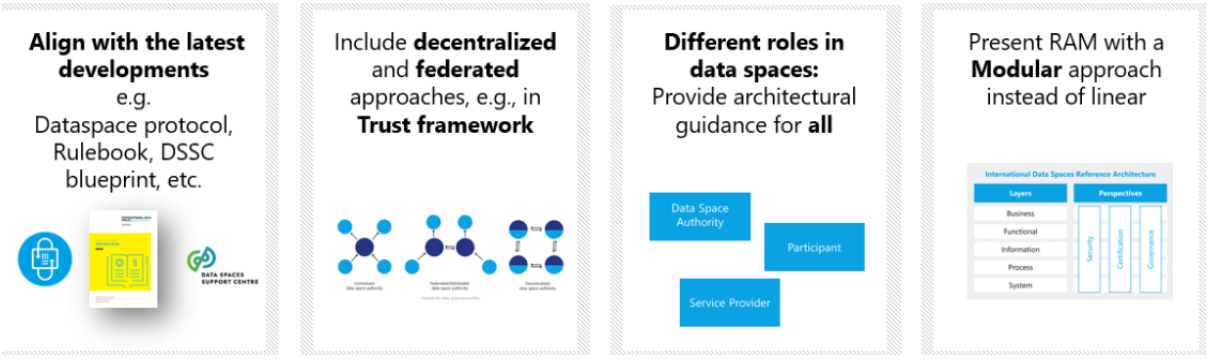

Figure 18. Improvements envisioned for IDS RAM 5 at high-level, Ref. IDSA internal, Architecture Working Group presentation [7].

It will provide an overview for technical readers on how to create an architecture for a data space, to participate in a data space, and to provide value added services for data spaces. To do so, RAM5 will sketch architectural decision areas for different roles in data spaces.

The RAM 5 document will not be a linear document like RAM 4 but will contain links between parts of the layers and perspectives. Other improvements in RAM 5 will include description of decentralized approaches and updates to the information model, among other things.

Timeline: The expected timeline is to provide a first draft of RAM 5 until the end of Quarter 2 2024 and a final document until the end of Quarter 2 2025.

IDSA has strengthened collaborations with key initiatives like Catena-X, which focuses on the automotive sector, and other industry-specific data spaces such as Prometheus-X. These partnerships are essential for creating sector-specific data ecosystems with tailored governance models (International Data Spaces).

The Prometheus-X project integrates an AI component in their Personal Data space ecosystem. They are producing coherent approaches to connect personal data store, various data and services providers and AI personal assistants via this connector to create new tools while ensuring data trust and sovereignty.

Specifically, they are trying to expand the dataspace infrastructure by working on:

- **Promoting a personal approach to data management by providing a standardized framework in compliance with the Data Spaces Protocol.** Thus, seamless integration across various platforms is accomplished by developing a centralized platform where educational data are stored and connected to the school platforms across the Data Space by using Personal Data Intermediary technology.
- **Secure and compliant data sharing and exchange.** The use of Protocols guarantees that data sharing is governed by consolidated contract policies and standardized metadata systems.

[7] https://internationaldataspaces.org/understanding-the-idsa-reference-architecture-model/

Using the Consent/Contracts Negotiation Agent will be an advancement in Dataspace ecosystem by handling contracts between users and organizations and automating responses to consent requests.

- **Automated matching and recommendations**: By utilizing advanced algorithms, the agent finds the finest services for users and the most suitable contracts for companies based on the preferences that have been expressed and the data-sharing regulations that are currently in place.
- **Decentralized AI training:** This building block provides a layer of contact between data providers and AI providers through secure data exchange. The incoming data give a new update in the weights of the models, and the final trained model can be extracted by the AI provider.

In that direction, Loria, an IDSA research institute partner, suggests some corrections to the federated learning direction which is adopted in the big data world by:

- **Developing indicators to deal with data heterogeneity** because each data provider's contribution in the generation of the final model is disproportionate [72].
- **Synchronizing the different devices** to produce Machine learning algorithms that are well-fitted for the training process throughout a wide range of heterogeneous Edge computing devices.

8.2 Gaia-X

As mentioned in previous points, the Gaia-X framework is currently being developed and enhanced to make it more robust and secure.

Talking about technical improvements, the launch of the new version of Gaia-X, called **Loire**, is scheduled for July 2024, and it will replace the current Tagus version. This new version is designed to address several key areas for improvement to ensure the framework's robustness and security. The enhancements and new features included in the Loire version aim to advance the following points:

- **More decentralization using blockchain techniques**: By leveraging blockchain technologies, the Gaia-X framework will achieve greater decentralization. This will enhance data integrity and security by distributing control across a wider network of nodes, reducing the risk of central points of failure and ensuring a more resilient infrastructure.
- **Rework of shapes/JSON-LD contexts with LinkML**: The update will include a comprehensive reworking of shapes and JSON-LD contexts using LinkML. This will provide more flexible and powerful data modeling capabilities, making it easier to define, validate, and manage complex data structures within the Gaia-X ecosystem.
- **Remote attestation (TPM/TEE, binary signature)**: The new version will incorporate remote attestation mechanisms, including Trusted Platform Module (TPM) and Trusted Execution Environment (TEE) technologies, along with binary signature verification. This will enhance the framework's ability to verify the integrity and authenticity of hardware and software components, bolstering security against tampering and unauthorized modifications.
- **Catalogues index implementation (transparency, veracity)**: The implementation of catalogues index aims to improve transparency and veracity within the Gaia-X framework. This will involve creating comprehensive indexes of available services and resources, ensuring that all data is accurately represented and easily accessible, thereby fostering trust and reliability in the ecosystem.

- **Policy reasoning (Using ODRL and credential evaluation)**: The Loire version will enhance policy reasoning capabilities by utilizing the Open Digital Rights Language (ODRL) and credential evaluation techniques. This will allow for more sophisticated and flexible policy enforcement, ensuring that access and usage policies are consistently applied and evaluated across the network.
- **Testbed for custom notary implementations**: Finally, the update will include a testbed environment for custom notary implementations. This will provide a controlled setting for developers to create, test, and refine their own notary solutions, facilitating innovation and ensuring that the Gaia-X framework can accommodate a wide range of notarization needs and use cases.

Additionally, new marketing enhancements are planned with the launch of the **Gaia-X Academy** in the final quarter of 2024. This initiative aims to make education on the Gaia-X project widely accessible. The academy will offer courses designed for various expertise levels, providing both foundational knowledge and advanced technical insights.

The curriculum will range from basic principles of Gaia-X to more complex topics, addressing areas such as the data economy, framework architecture, and technical components. Participants will develop the skills needed to lead related projects and gain the technical expertise required to facilitate Gaia-X adoption. Furthermore, the academy will provide an in-depth analysis of the Gaia-X architecture, equipping participants with the knowledge to design compliant solutions and effectively support business objectives.

The following is the course overview:

- Gaia-X Fundamentals
- Gaia-X Trust Framework Overview
- Gaia-X Trust Framework Technical Focus
- Catalogue Browsing Overview
- Catalogue Browsing Technical Focus
- Contracts and Policy Reasoning Overview
- Policy Reasoning Technical Focus
- Gaia-X Compliance and Label Document
- Gaia-X Architecture Document
- How to deploy a Gaia-X Digital Clearing House
- Gaia-X Certificates - Functional Knowledge Level 1

8.3 FIWARE

The FIWARE platform has undergone substantial advancements, particularly in the domains of NGSI-LD and data usage control, which significantly enhance its functionality and applicability in the realm of smart solutions.

- NGSI-LD, or Next Generation Service Interface Linked Data, represents a transformative evolution of the original NGSI context interfaces. Initially standardized by the Open Mobile Alliance, NGSI has been enhanced within the FIWARE ecosystem to improve the management of context information. The new version not only supports more complex data models but also promotes

semantic interoperability, making it a key player for innovative smart solutions across various sectors.

- In addition to the enhancements in context management, the incorporation of data usage control mechanism within the FIWARE framework marks a critical advancement. This feature allows organizations to have precise control over who can access data and how it can be used. By setting clear policies, organizations can ensure their data management practices comply with regulations while also aligning with their business goals. This level of granularity in data governance is essential for fostering secure and compliant data exchanges, which helps build trust among stakeholders.

Additionally, the recent introduction of the FIWARE Data Space Connector [73] marks a significant development within the FIWARE ecosystem, catering to organizations that need a robust connector for data spaces. This connector is designed to enable secure and seamless data sharing across different platforms, in alignment with the Data Spaces Business Alliance (DSBA) technical framework and the latest EU DigitalID standards. It supports various APIs, including NGSI-LD and NGSIv2, and is expected to expand further by adding support for IDS protocols and TM Forum APIs. This expansion is critical for enhancing the flexibility and interoperability of data ecosystems, thus allowing organizations to engage effectively in trusted, large-scale data spaces.

Another collaboration currently ongoing is with Latitudo 40 [74], which is working to integrate its datasets into the FIWARE Marketplace. This marketplace is a vital part of the FIWARE ecosystem, offering a wide range of ready-to-use solutions, services, and data models that facilitate the development of smart applications. By using the marketplace, companies can find verified solutions that make it easier to build smart, sustainable cities and tackle the complex challenges of today's urban environments.

These developments not only improve the functionality of the platform but also address the growing need for secure and compliant data practices in an increasingly interconnected and data-driven world.https://www.fiware.org/2023/12/28/paving-the-way-for-a-digital-future-fiwares-great-success-of-the-year-2023/https://www.fiware.org/2024/09/03/data-spaces-and-digital-twins-for-empowering-cities-to-embrace-sustainability/

8.4 IHAN

As mentioned in previous points, the IHAN data economy architecture was defined to allow companies, governments and individuals to share data in an easy and trusted manner with users' consents.

Although the project in which this framework was developed ended in 2021, it is possible to identify different advancements aligned with the European data strategy:

- **Strengthening data sovereignty:** IHAN enables individuals and organisations to have greater control over their data, deciding how, when and with whom it is shared. This aligns with the European data strategy's goal of empowering European citizens and businesses in the digital environment.
- **Creating a single market for data:** The interoperability and common standards promoted by IHAN facilitate the exchange of data across different sectors and countries in the EU, an essential element in building a single market for data.
- **Fostering innovation:** By facilitating access to and sharing of data, IHAN can stimulate innovation in Europe, enabling the development of new data-driven services and business models.
- **Increased trust in the data ecosystem:** IHAN's focus on transparency, security and user control can contribute to increasing trust in the European data ecosystem.

9 Set of extensions requirements to existing data spaces architectures that address the identified interoperability, security and privacy gaps.

In response to the gaps identified in **Chapter 7**, this section outlines the necessary interoperability enhancements that should be implemented in the data spaces architectures discussed in **Chapter 3**. The goal is to enhance the capability of data spaces to efficiently handle diverse technical, organizational, and legal challenges, ensuring seamless, secure, and trustworthy data sharing.

9.1 Technical Interoperability Enhancements

The technical challenges identified in Chapter 7 include the **lack of protocol standardization**, **data format compatibility issues**, and **identity management difficulties**. These must be addressed to ensure smooth data exchange between systems. The following enhancements are recommended [77]:

- **Standardized Communication Protocols (TI1)**: All data space architectures should implement a common set of communication protocols, ensuring compatibility across different systems. For example, Gaia-X and IDS-RAM emphasize standard protocols such as **Dataspace Protocol (DSP)** and **NGSI-LD API** for secure data exchange but they are not interoperable with the other architecture. A common secure communication protocol should be selected and extended across all architectures at least for inter architecture communication.
- **Unified Data Formats (TI2)**: Interoperability suffers from varied data formats in different systems. The adoption of common data models (e.g., **Smart Data Models** in FIWARE) or transformation services should be integrated across data spaces to ensure that data can be exchanged seamlessly.
- **Standardized Identity and Access Management (TI3)**: Across all architectures providing authorization mechanisms for data access to all the users across architectures.
- **Enhanced Connector Technology (TI4)**: To facilitate inter-data space communication, all architectures must support flexible connectors that can integrate with multiple platforms and adapt to legal requirements.

9.2 Semantic Interoperability Enhancements

Semantic interoperability is crucial for the effective exchange of data with preserved meaning. **Vocabulary misalignment (SI1)** and **contextual inconsistencies (SI2)** were major challenges identified in Chapter 7. The following improvements are essential:

- **Ontology and Vocabulary Alignment (SI1)**: All architectures must adopt or create shared vocabularies, allowing for cross-domain data sharing. **IDS-RAM's Common Information Model** or the **FIWARE Smart Data Models** should serve as a blueprint for creating a unified data structure across different systems.
- **Context Preservation (SI2)**: Implementations must include mechanisms to retain the context of shared data across systems, ensuring that meaning is not lost during exchange. The **NGSI-LD API** in FIWARE and the **Data Space Protocol (DSP)** in IDS-RAM provide context-awareness during data transmission.
- **Cross-Domain Semantics and Interoperability (SI3-SI6)**: Seamless interoperability across multiple domains is crucial because it enables diverse systems, organizations, and industries to communicate and exchange data efficiently without barriers. To this end, data spaces architectures must support a robust and flexible semantic framework that capture cross-domain knowledge representation.

- **Metadata Standardization (SI4)**: Data space architectures must implement standardized metadata semantics to ensure consistent interpretation and use across diverse datasets. Standardized metadata is essential for achieving cross-domain interoperability, as it provides a common understanding of data attributes, allowing systems from different domains to effectively communicate and exchange information. Additionally, standardized metadata enables the development of data conversion tools that can seamlessly translate between different data formats, further enhancing interoperability and reducing complexity in integrating data from various sources.

9.3 Organizational Interoperability Enhancements

Governance model alignment and **participant onboarding** are key areas needing enhancement. These improvements are crucial for the smooth integration of new participants and the maintenance of trust within the data ecosystem:

- **Governance Alignment (OI1)**: A common governance framework that includes shared principles for data management, privacy, and security is needed across all data spaces. **Gaia-X** provides a robust governance model emphasizing **European values of data sovereignty and GDPR compliance**, which could be extended to other architectures.
- **Participant Onboarding and Compliance (OI2)**: Onboarding new participants is complex, particularly in decentralized data spaces. Implementing standardized onboarding procedures, like the **Gaia-X Registry**, can simplify compliance checks and ensure that all participants meet minimum security and privacy standards.
- **Data Sharing Agreements (OI3)**: A consistent data sharing approach is critical because, innocent or unclear sharing agreements can create legal disputes especially when operating across different jurisdiction. To this end, a standardized framework for data sharing agreement must be established to build trust between participants, ensure accountability, and foster collaboration while maintaining data integrity and security.

9.4 Legal and Regulatory Interoperability Enhancements

Legal and regulatory challenges, particularly around **GDPR compliance (LR2)** and **data sharing agreements (LR3)**, were highlighted in Chapter 7. These must be addressed to enable seamless cross-border data sharing:

- **Legal Framework Alignment (LR1, LR4):** A carefully designed federated Data Space structure can address the different legal requirements across sectors and countries.
- **GDPR Compliance (LR2)**: All architectures should integrate **automated auditing tools** for monitoring data access and ensuring compliance with GDPR. Gaia-X, which already incorporates **lineage tracking** for accountability, should be a model for other systems.
- **Data Sharing Agreements (LR3)**: A standardized legal framework for data-sharing agreements should be adopted across all data spaces to facilitate trust and legal consistency. This can be modelled on the **Usage Control mechanisms** in IDS-RAM, which enforce data provider conditions, and Gaia-X's **ODRL-based policy reasoning**.

9.5 Scalability and Flexibility

To future-proof data space architectures, it is essential to address scalability and flexibility, particularly as data volumes increase:

- **Federated Architecture for Scalability**: A federated and decentralized architecture, such as **Gaia-X's federated model**, should be implemented across data spaces to ensure scalability without

compromising performance. Distributed computing technologies, such as **edge computing**, should be incorporated to handle large datasets efficiently.

- **Modular Architecture for Flexibility**: Data spaces must adopt a **modular architecture** that allows the addition or removal of components based on the needs of the ecosystem. **FIWARE's modular approach** enables flexibility and allows for the seamless integration of new technologies, ensuring that the architecture remains adaptable to evolving requirements.

10 Conclusions

This document provides a complete and detailed vision of European data spaces architectures and other data-sharing initiatives. These data spaces are instrumental in driving the digital economy, as they enable secure and efficient data exchange across organizations, industries, and regions. Through the use of a structured methodology, it has been possible to assess the strengths and limitations of each framework in terms of functionality, scalability, interoperability, and security.

IDS-RAM stands out for its robust approach to secure data exchange, providing a standardized framework that emphasizes data sovereignty and trust between participants. Its framework is foundational to many other data sharing platforms, setting global standards for secure and controlled data exchange. The FIWARE platform, on the other hand, offers an open-source approach, with a strong emphasis on smart cities, digital twins, and IoT integration. FIWARE's flexibility and scalability make it highly adaptable to a wide range of use cases, yet ensuring its seamless integration with other frameworks like IHAN remains a to be addressed for greater adoption.

Gaia-X, one of the most ambitious European data space projects, focuses on creating a federated data infrastructure that promotes European data sovereignty. However, Gaia-X's complexity and reliance on multistakeholder collaboration can create barriers to rapid implementation, especially when it comes to balancing the diverse needs of industries and regions. IHAN takes a more human-centric approach, aiming to empower individuals to control their personal data. IHAN offers a vision of data spaces where data portability and ethical data use are at the forefront, but its broader adoption faces challenges as it seeks alignment with larger, more industry-focused frameworks like IDS-RAM and Gaia-X.

The exploration of additional data-sharing initiatives, like MyData, BDVA and DKSR, further underscores the importance of creating interconnected environments where data can flow securely and efficiently. These initiatives highlight the role of government and private sector collaboration in ensuring that data can be shared while respecting privacy, ethical use, and security standards. MyData contributes significantly by promoting human-centric data governance, which aligns closely with ethical data sharing. Similarly, BDVA plays a critical role in advancing innovation in data driven technologies, particularly by fostering cooperation between industry and research. The Data Sharing Coalition adds value by fostering secure, cross-sectoral data exchange, yet it operates on a more practical, use-case-driven level compared to the comprehensive structures offered by IDS-RAM, FIWARE, Gaia-X, and IHAN.

Throughout the analysis of these data spaces, security and privacy have emerged as crucial issues. Identifying vulnerabilities related to data misuse, breaches, and privacy violations is essential to improving trust and adoption rates. While frameworks like IDS-RAM and Gaia-X offer strong security architectures aimed at protecting data integrity and privacy, gaps remain, particularly in areas of cross-platform interoperability and ensuring end-to-end data protection. Strengthening these measures is crucial to maintaining data sovereignty and preventing unauthorized access or misuse of shared data.

The proposed set of extension requirements and enhancements outlined in this document address these concerns, focusing on improving governance structures, ensuring interoperability, and establishing robust security protocols. By implementing standardized practices and safeguarding data sovereignty, these recommendations aim to strengthen existing frameworks, enabling the creation of secure, trustworthy, and future-proof data sharing environments.

In conclusion, the future success of data spaces lies in the ability to promote collaboration while safeguarding privacy and security. The study done in this document provides a strategic direction for developing data spaces, ensuring that the digital economy can grow in a secure, interoperable, and privacy-aware environment.

11 References

[1] https://internationaldataspaces.org/understanding-the-idsa-reference-architecture-model/

[2] Juan J MONTERO-PASCUAL, Matthias Finger, and Francisco Miguel DE ABREU DUARTE. Creating a common European mobility data space. 2023.

[3] Singh, M., & Jain, S. (2011). A Survey on Dataspace.

[4] Christos Doulkeridis, Georgios M Santipantakis, Nikolaos Koutroumanis, George Makridis, Vasilis Koukos, George S Theodoropoulos, Yannis Theodoridis, Dimosthenis Kyriazis, Pavlos Kranas, Diego Burgos, et al. Mobispaces: An architecture for energy-efficient data spaces for mobility data. In 2023 IEEE International Conference on Big Data (BigData), pages 1487–1494. IEEE, 2023.

[5] Nagel L., Lycklama D. (2021): Design Principles for Data Spaces. Position Paper. Version 1.0. Berlin

[6] Hummel, P., Braun, M., Tretter, M., & Dabrock, P. (2021). Data sovereignty: A review. Big Data & Society, 8(1), 2053951720982012.

[7] Steinbuss, S. et al., Semantic Interoperability in Data Spaces, International Data Spaces Association, 2024 https://doi.org/10.5281/zenodo.10964377

[8] Holger Drees, Dennis O Kubitza, Johannes Lipp, Sebastian Pretzsch, and Christoph Schlueter Langdon. Mobility data space–first implementation and business opportunities. In ITS World Congress, 2021.

[9] Flavio Cirillo, Gürkan Solmaz, Everton Luís Berz, Martin Bauer, Bin Cheng, and Ernoe Kovacs. A standard-based open source iot platform: Fiware. IEEE Internet of Things Magazine, 2(3):12–18, 2019

[10] https://internationaldataspaces.org/wp-content/uploads/dlm_uploads/IDSA-Position-Paper-Semantic-Interoperability-in-Data-Spaces-1.pdf

[11] https://internationaldataspaces.org/wp-content/uploads/Reflections-on-the-DGA-and-Data-Intermediaries.pdf

[12] https://docs.internationaldataspaces.org/ids-knowledgebase/dataspace-protocol/transfer-process/transfer.process.protocol#2-message-types

[13] https://docs.internationaldataspaces.org/ids-knowledgebase/dataspace-protocol/overview/terminology#dataset

[14] https://docs.internationaldataspaces.org/ids-knowledgebase/dataspace-protocol/overview/terminology#provider

[15] https://docs.internationaldataspaces.org/ids-knowledgebase/dataspace-protocol/overview/terminology#consumer

[16] https://gaia-x.eu/wp-content/uploads/2022/06/Gaia-x-Architecture-Document-22.04-Release.pdf

[17] https://gaia-x.eu/wp-content/uploads/2022/05/Gaia-X_Policy-Rules_Document_v22.04_Final.pdf

[18] https://docs.gaia-x.eu/framework/?tab=clearing-house

[19] Ahle, U., & Hierro, J. J. (2022). FIWARE for Data Spaces. En B. Otto, M. ten Hompel, & S. Wrobel (Eds.), Designing Data Spaces: The Ecosystem Approach to Competitive Advantage (pp. 395-417). Springer International Publishing. https://doi.org/10.1007/978-3-030-93975-5_24

[20] Jeong, S., Kim, S., & Kim, J. (2020). City Data Hub: Implementation of Standard-Based Smart City Data Platform for Interoperability. Sensors, 20(23), Article 23. https://doi.org/10.3390/s20237000

[21] Bauer, M. (2022). FIWARE: Standard-based Open Source Components for Cross-Domain IoT Platforms. 2022 IEEE 8th World Forum on Internet of Things (WF-IoT), 1-6. https://doi.org/10.1109/WF-IoT54382.2022.10152259

[22] https://www.fiware.org/news/interoperability-of-fiware-and-gs1-standards-boosts-innovation-in-the-iot-space/

[23] Pozo, A., Alonso, Á., & Salvachúa, J. (2020). Evaluation of an IoT Application-Scoped Access Control Model over a Publish/Subscribe Architecture Based on FIWARE. Sensors, 20(15), Article 15. https://doi.org/10.3390/s20154341

[24] https://fiware-legal.readthedocs.io/en/latest/PersonalDataProtectionPolicy.html

[25] De Nardis, L., Mohammadpour, A., Caso, G., Ali, U., & Di Benedetto, M.-G. (2022). Internet of Things Platforms for Academic Research and Development: A Critical Review. Applied Sciences, 12(4), Article 4. https://doi.org/10.3390/app12042172

[26] https://fiware.github.io/specifications/ngsiv2/stable/

[27] https://github.com/FIWARE/catalogue/releases

[28] https://www.fiware.org/news/fiware-8-4-1-is-out/

[29] https://www.sitra.fi/en/projects/ihan-pilot-projects/#what-is-it-about

[30] https://sales.sfs.fi/en/index/tuotteet/SFS/CEN/ID5/1/996254.html.stx?_ga=2.259869898.1588871291.1639738429-17692516.1639738429

[31] https://www.youtube.com/watch?v=ZNwWmUU134GA

[32] https://www.youtube.com/watch?v=TfZWRGN8cXU

[33] https://www.youtube.com/watch?v=E29X3UD8zBw

[34] https://dssc.eu/space/BVE/357073006/Data+Spaces+Blueprint+v1.0

[35] Alon Halevy, Michael Franklin, and David Maier. Principles of dataspace systems. In Proceedings of the twenty-fifth ACM SIGMOD-SIGACT- SIGART symposium on Principles of database systems, pages 1–9, 2006.

[36] DSBA – Technical Convergence Discussion Document https://data-spaces-business-alliance.eu/dsba-releases-technical-convergence-discussion-document/

[37] Chris Schlueter Langdon and Riyaz Sikora. Creating a data factory for data products. In Smart Business: Technology and Data Enabled Innovative Business Models and Practices: 18th Workshop on e-Business, WeB 2019, Munich, Germany, December 14, 2019, Revised Selected Papers 18, pages 43–55. Springer, 2020.

[38] Grothe, M. (s. f.). Exploring data space initiatives.

[39] https://bdva.eu/about/

[40] https://european-big-data-value-forum.eu/2024-edition/About/

[41] https://www.dksr.city/en/the-dataplatform/

[42] Siska, V., Lorünser, T., Krenn, S., & Fabianek, C. (2024). Integrating Secure Multiparty Computation into Data Spaces: Proceedings of the 14th International Conference on Cloud Computing and Services Science, 346-357. https://doi.org/10.5220/0012734600003711

[43] Thomás Oliveira, C., Moreira, R., de Oliveira Silva, F., Sanches Miani, R., & Frosi Rosa, P. (2018). Improving Security on IoT Applications Based on the FIWARE Platform. *2018 IEEE 32nd International Conference on Advanced Information Networking and Applications (AINA)*, 686-693. https://doi.org/10.1109/AINA.2018.00104

[44] Paulo H Rettore, Guilherme Maia, Leandro A Villas, and Antonio AF Loureiro. Vehicular data space: The data point of view. IEEE Communications Surveys & Tutorials, 21(3):2392–2418, 2019.

[45] Perata, J. P., & Betarte, G. (2023). A Security Analysis of a Referential Architecture of the FIWARE Platform. 2023 XLIX Latin American Computer Conference (CLEI), 1-9. https://doi.org/10.1109/CLEI60451.2023.10346164

[46] https://www.privacy-regulation.eu/en/recital-71-GDPR.htm

[47] https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence

[48] Truong, N., Sun, K., Wang, S., Guitton, F., & Guo, Y. (2021). Privacy preservation in federated learning: An insightful survey from the GDPR perspective. Computers & Security, 110, 102402. https://doi.org/10.1016/j.cose.2021.102402

[49] https://secureprivacy.ai/blog/navigating-data-privacy-2024

[50] Lachner, C., Rausch, T., & Dustdar, S. (2021). A Privacy Preserving System for AI-assisted Video Analytics. 2021 IEEE 5th International Conference on Fog and Edge Computing (ICFEC), 74-78. https://doi.org/10.1109/ICFEC51620.2021.00018

[51] https://gdpr-info.eu/art-8-gdpr/

[52] Edward Curry and Edward Curry. Dataspaces: fundamentals, principles, and techniques. Real-time Linked Dataspaces: Enabling Data Ecosystems for Intelligent Systems, pages 45–62, 2020.

[53] Pierfrancesco Bellini, Stefano Bilotta, Enrico Collini, Marco Fanfani, and Paolo Nesi. Data sources and models for integrated mobility and transport solutions. Sensors, 24(2):441, 2024.

[54] https://gdpr-info.eu/art-12-gdpr/

[55] https://ec.europa.eu/isa2/sites/default/files/eif_brochure_final.pdf

[56] https://standards.iso.org/ittf/PubliclyAvailableStandards/c066639_ISO_IEC_19941_2017.zip

[57] https://docs.internationaldataspaces.org/ids-knowledgebase/v/idsa-rulebook/idsa-rulebook/3_interoperability

[58] Alberto Dognini, Chandra Challagonda, Erik Maqueda Moro, Kristian Helmholt, Henrik Madsen, Laura Daniele, Laurent Schmitt, Lynda Temal, Olivier Genest, Philippe Calvez, et al. Data spaces for energy, home and mobility. 2022.

[59] https://internationaldataspaces.org/offers/dataspace-protocol/

[60] Joyce, A., & Javidroozi, V. (2024). Smart city development: Data sharing vs. data protection legislations. Cities, 148, 104859. https://doi.org/10.1016/j.cities.2024.104859

[61] Bastiaansen, H. J. M., Kollenstart, M., Dalmolen, S., & Engers, T. M. van. (2020). User-centric network-model for data control with interoperable legal data sharing artefacts: Improved data sovereignty, trust and security for enhanced adoption in interorganizational and supply chain is applications. 24th Pacific Asia Conference on Information Systems: Information Systems (IS) for the Future, PACIS 2020. https://research.utwente.nl/en/publications/user-centric-network-model-for-data-control-with-interoperable-le

[62] https://datacellarproject.eu/uncategorized/semantic-interoperability-in-data-spaces/

[63] https://www.fiware.org/smart-data-models/

[64] https://www.baidata.eu/en/dssc-insights-series--standardisation-in-data-spaces

[65] https://www.privacy-regulation.eu/en/32.htm

[66] Conley, E., & Pocs, M. (2018). GDPR Compliance Challenges for Interoperable Health Information Exchanges (HIEs) and Trustworthy Research Environments (TREs). European Journal for Biomedical Informatics, 14(3). https://doi.org/10.24105/ejbi.2018.14.3.7

[67] Otto, B., Ten Hompel, M., & Wrobel, S. (Eds.). (2022). Designing Data Spaces: The Ecosystem Approach to Competitive Advantage. Springer International Publishing. https://doi.org/10.1007/978-3-030-93975-5

[68] Zarko, I., Soursos, S., Gojmerac, I., Ostermann, E., Insolvibile, G., Plociennik, M., Reichl, P., & Bianchi, G. (2017). Towards an IoT Framework for Semantic and Organizational Interoperability. https://doi.org/10.1109/GIOTS.2017.8016253

[69] Boukhers, Z., Lange, C., & Beyan, O. (2023). Enhancing Data Space Semantic Interoperability through Machine Learning: A Visionary Perspective. Companion Proceedings of the ACM Web Conference 2023, 1462-1467. https://doi.org/10.1145/3543873.3587658

[70] Rahman, H., & Hussain, Md. I. (2020). A Comprehensive Survey on Semantic Interoperability for Internet of Things: State-of-the-Art and Research Challenges. Transactions on Emerging Telecommunications Technologies, 31. https://doi.org/10.1002/ETT.3902

[71] Sousa, P. R., Magalhães, L., Resende, J. S., Martins, R., & Antunes, L. (2021). Provisioning, Authentication and Secure Communications for IoT Devices on FIWARE. Sensors (Basel, Switzerland), 21(17), 5898. https://doi.org/10.3390/s21175898

[72] Lal, N., Qamar, S., & Shiwani, S. (2016). Search Ranking for Heterogeneous Data over Dataspace. Indian journal of science and technology, 9.

[73] https://www.fiware.org/2023/12/28/paving-the-way-for-a-digital-future-fiwares-great-success-of-the-year-2023/

[74] https://www.fiware.org/2024/09/03/data-spaces-and-digital-twins-for-empowering-cities-to-embrace-sustainability/

[75] IDSA Rulebook, IDSA website. https://docs.internationaldataspaces.org/ids-knowledgebase/idsa-rulebook

[76]The General Data Protection Regulation (GDPR), Regulation (EU) 2016/679 of the European Parliament

[77] Pettenpohl, H., Spiekermann, M., & Both, J.R. (2022). International Data Spaces in a Nutshell. Designing Data Spaces.

[78] IDSA Reference Architecture Model4.0. IDSA Github: https://github.com/International-Data-Spaces-Association/IDS-RAM_4_0