INTERNATIONAL DATA SPACES ASSOCIATION

Position Paper | Version 1.0 | April 2024

Semantic Interoperability in Data Spaces

- ● Position Paper of members of the IDS Association
- ○ Position Paper of bodies of the IDS Association
- ○ Position Paper of the IDS Association
- ○ White Paper of the IDS Association

Publisher

International Data Spaces Association
Emil-Figge-Str. 80
44227 Dortmund
Germany

Editor

Sebastian Steinbuss, IDSA

Authors & Contributors

Wouter van den Berg, TNO
Sthefan Gabriel Berwanger, INESTEC
Sonia Bilbao Arechabala, tecnalia
Izaskun Fernandez, Tekniker
Gonzalo Gil Inchaurza Phd, Tekniker
Pieter Klinker, TNO
Viivi Lähteenoja, 1001 lakes
Andreas Müller, Schaeffler
Michiel Stornebrink, TNO
Sebastian Steinbuss, IDSA
Anil Turkmayali, IDSA

Cite as

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

Contributing projects

DATES European Data Space for Tourism
demeter
DigiChecks
DIVINE

Contents

1 Introduction

Data sovereignty is a central aspect of the International Data Spaces. It can be defined as a natural person's or legal entity's “supreme authority with regard to the digital domain particular to themselves“ [1], [2]. The International Data Spaces initiative proposes a Reference Architecture Model for this capability and related aspects, including requirements for secure and trusted data exchange in business ecosystems.

In addition to data sovereignty, International Data Spaces aims at meeting the following strategic requirements to build a reliable foundation for trusted data sharing in Data Spaces.

Figure 1 Functional Requirements (source: IDS RAM 4)

The IDSA Rulebook provides a clear guideline for the mandatory and optional requirements of Data Spaces. Vocabularies and semantic models are identified as part of the mandatory foundation of Data Spaces.

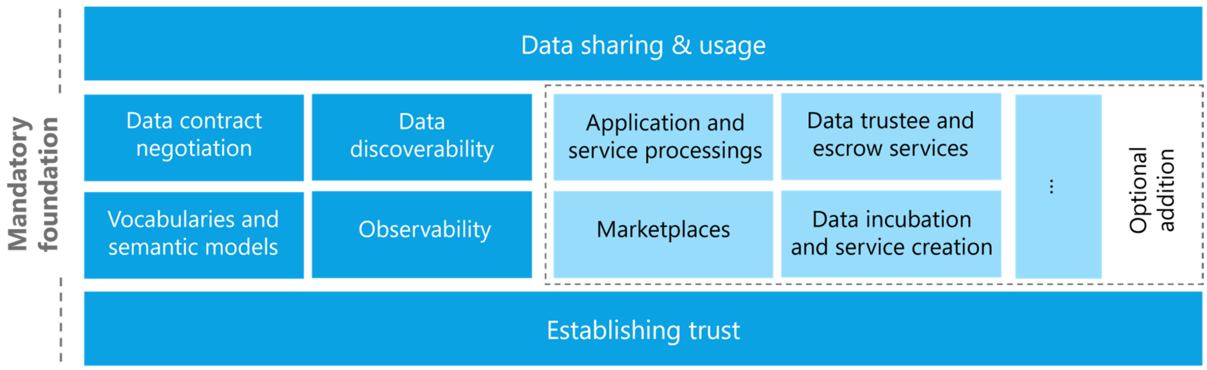

Figure 2 Foundational concepts of Data Spaces (source: IDSA Rulebook)

This paper focuses on the semantic interoperability aspects of data spaces, as interoperability is a key concern in data spaces because, simply put, nothing works without it. Further, interoperability can be addressed on multiple levels. The New European Interoperability Framework [3] defines an interoperability model with four layers of interoperability: technical, semantic, organizational and legal.

Technical and semantic interoperability are covered by data connectors, which are defined in the IDS-RAM [2] and described in the IDSA Data Connector Report [4]. As described in the remainder of the document, both may require interaction with additional services or components.

Legal interoperability and operational interoperability can be achieved by the Policies and Rules of a specific data space instance and are typically managed by a data space authority. More information can be found in the IDSA Rulebook[5].

In addition to the new European Interoperability Framework [3], which is applicable to all digital public services, ISO/IEC 21823-1:2019 [6] introduces a five-facet model specifically for IoT systems interoperability: transport, syntactic, semantic, behavioral and policy interoperability. Although they use slightly different names, both frameworks address very similar concepts.

Technical interoperability in the EIF Framework covers both the transport and syntactic interoperability of ISO/IEC 21823-1. On the one hand, transport interoperability [7] is responsible for guaranteeing the communication and error-free delivery of data between different entities, which may be connected to different networks. Quality of Service (QoS) requirements such as timeliness, ordering, durability, and lifespan are considered in this facet. On the other hand, syntactic interoperability enables the formats of the exchanged information to be understood by the participating systems. Aspects of technical interoperability include interface specifications, interconnection services, data integration services, data presentation and exchange, and secure communication protocols.

Semantic interoperability enables the exchange of data between entities using understood data information models (or semantic meanings) [8]. According to [3], [9], semantic interoperability is achieved when interacting systems attribute the same meaning to an exchanged piece of data, ensuring consistency of the data across systems regardless of individual data format. This consistency of meaning can be derived from pre-existing standards or agreements on the format and meaning of data or it can be derived in a dynamic way using shared vocabularies either in a schema form and/or in an ontology driven approach.

So, in short, transport interoperability deals with data delivery (i.e. sending the data); syntactic interoperability allows reading the data in a known format and grammar; whereas semantic interoperability is responsible for the meaning, enabling the unambiguous interpretation and understanding of data.

According to EIF, organizational interoperability refers to the way in which public administrations align their business processes, responsibilities, and expectations to achieve commonly agreed and mutually beneficial goals. It also aims to meet the requirements of the user community by making services available, easily identifiable, accessible and user focused. In this layer, the relationship between service providers and service consumers must be clearly defined. Similarly, the behavioral interoperability defined in ISO/IEC 21823-1:2019 ensures that the actual result exchanged achieves the expected outcome. Table 1provides a summary of the EIF and ISO/IEC 21823-1:2019.

Finally, legal or policy interoperability is about ensuring that organizations operating under different legal frameworks, policies and strategies can work together.

Table 1 Comparison EIF and ISO/IEC 21823-1:2019

European Interoperability Framework	ISO/IEC 21823-1:2019	What it means
Technical interoperability	Transport interoperability	Deals with data delivery
	Syntactic interoperability	Allows reading the data in a known format and grammar
Semantic interoperability	Semantic interoperability	Responsible for the meaning, enabling the unambiguous interpretation and understanding of data
Organizational interoperability	Behavioral interoperability	Refers to the way in which business processes, responsibilities and expectations are aligned to achieve commonly agreed and mutually beneficial goals
Legal interoperability	Policy interoperability	Ensures that organizations operating under different legal frameworks, policies and strategies can work together

This document provides an overview of means for and examples of specifically semantic interoperability.

1.1 Need for specifically semantic interoperability in data spaces

Clarity about the meaning of data is essential to ensure that data can be accurately and consistently interpreted and used by different people and systems. When the meaning of shared data is unclear, it can lead to miscommunication and misinterpretation, resulting in errors and poor decisions. Organizations that find themselves dealing with ambiguous data therefore spend a lot of effort in mapping that data to the formats and structures that their IT systems expect. Since this is time-consuming and costly, the lack of shared meaning of data is a major barrier to data sharing and therefore to the realization of the Digital Single Market strategy.

The ability of IT systems to exchange data with unambiguous, shared meaning is called semantic interoperability. Semantic interoperability is an essential requirement for federated data networks such as IDS and data spaces in general. It requires that data providers and data consumers in the network express their data offering or need using explicit reference to a common vocabulary. The commonality is important; if data is provided with references to a vocabulary that the receiving party is unfamiliar with, then the need to spend integration effort on their end remains. Vocabularies become common (shared) through the process of standardization (using different means).

Many industries and other business ecosystems have turned to **open standardization** to achieve semantic interoperability between their members. Open standardization means that members collaboratively maintain and develop semantic standards. It is a continuous balancing act between the need for strict uniformity to keep data consistent and easy to understand, and the need to accommodate for the fact that different organizations have different requirements for their data.

This means there is often a limit to the level of semantic interoperability that can be achieved. Every member in a business ecosystem operates with a different world view. These differences arise from operating in different jurisdictions, in different domains, carrying out different business processes, serving different markets, offering different services, and so on. High variety between members in business ecosystems means that any semantic standard for that community will have to allow for flexibility, which means some integration effort will remain necessary. Low variety allows for stricter semantic standards that bring more uniformity, thus allowing for more efficient data sharing and automation. [9]

In any case, governance must be put in place to make sure that a semantic standard serves the needs of the community as best as possible and will remain doing so. How these governance processes can be organized is discussed in the governance perspective of the IDS-RAM [2]. How the semantics of data can be made available and used is described in the process perspective of the IDS-RAM[2].

From a technological perspective, achieving semantic interoperability requires the use of semantic technologies to create Linked Data. Semantic technologies such as RDF, SHACL, OWL and SKOS allow us to enrich data with meaning by creating links to other datasets and vocabularies, enabling automatic reasoning over data through rules. The role of vocabularies and other semantic technologies in IDS is discussed in the Layers of the IDS Reference Architecture Model (section 3)[2]).

2 General approach to semantic interoperability

The rise of data spaces goes hand in hand with the demand for connecting, interpreting, and processing large bodies of heterogeneous data provided or consumed by a variety of interlinked systems in and across various domains. To enable these tasks to be performed efficiently, effectively, and ideally error-free, the data needs to be properly described by semantic models that capture the meanings of the data. This need forms a cornerstone for the automated discovery and utilization of data by consumers and providers. With data that is well-described, humans and systems alike are generally enabled to better understand and interoperate with each other. When introducing dynamicity in connecting the systems, i.e., having consumers dynamically interacting with providers, this emphasizes the requirement for continuous semantics and standards especially with view to end-to-end scenarios[10]. This holds true especially in data space settings and when forming chains of data processing, in which some of the providing systems involved might even be AI-based, which might not be known at the time of the consumer system's release. The semantics and semantic models to enable such complex connected use cases composed of both systems and data need to be carefully developed. This development path is outlined in [11] through the four basic scenarios of **Understand**, **Find**, **Update**, and **Operate**, which build on top of each other. These outline the challenges of

- **Understanding** the entities, i.e., systems, and data to be managed by providing models that permit proper interpretation of the data and the tasks involved in handling them.

- **Finding** the right, i.e., properly understood, data and models for the intended handling based on given criteria, which effectively requires the ability to perform queries.

- **Updating** found data and models based on given criteria to keep track of various changes influencing the entities and data.

- **Operating** on the updatable entities and data based on given criteria, i.e., performing operations of different complexities.

Mastering these steps allows for realizing many data-driven use cases that nowadays need considerable manual efforts for system integration quicker and more reliable.

However, even the most basic scenario of Understand is subject to a variety of heterogeneities that impair semantic interoperability. Starting from lossless transport of meanings of data between consumers and providers with different capabilities over a lack of commonly used and well-formalized standard models and their reuse up to the integrability and integration of such models, the challenges are complex.

Establishing semantic interoperability and the process thither can still prove beneficial in many ways. On a purely technical level this leverages a clearer understanding of the data, of the systems involved in processing the data, the processes the data and systems are subject to, and of the roles data and systems can play in the forming of end-to-end value chains. As a result, after the initial knowledge engineering phase, the efforts to build, operate and maintain solutions over longer lifetimes benefit from the well-described and stable knowledge as it permits to quickly connect with participants, i.e., systems or data sources.

On an organizational level it promotes harmonized and thus more efficient approaches to shaping use cases and to operating and managing data-driven solutions and data ecosystems, including the standards required in fueling them. This is due to the common meanings the involved development, operation and governance processes can be based on and can be aligned across different domains, organizations or even jurisdictions. Furthermore, it facilitates onboarding, training and coordination of the workforce entrusted with performing the processes but often unfamiliar with semantics.

To benefit from the advantages of full semantic interoperability, several challenges need to be mastered. Given that in data spaces not only data but also the systems involved in processing need to be understood and coupled to form end-to-end chains, the requirements here are manifold [11].

First, to fulfil the Understand scenario, there is the fundamental requirement for reusable and queryable models for both data and systems that need to be integrated with each other, especially with view to envisaged automated operations of systems. The latter imposes the additional need for automated matching and mapping data to the right models used for processing the data. As there are multitudes of domain-specific models and standards in place which may also follow different approaches in modelling, the resulting semantic heterogeneity needs remedial actions towards a continued but combined use of the established range of models.

With view to the capabilities needed to enable the Find scenario, unstructured data needs to be handled by proper models, which in turn need to be suitably integrated to support consumer-specific views on data and systems landscapes. Again, this needs promotion of sufficiently high degrees of formalization and the reuse of models. Here, mechanisms for ontology integration form a cornerstone.

The Update scenario then requires both data and models to be traceably modifiable during their lifecycles. This involves suitable mechanisms for executing model-based CRUD operations with the models themselves needing to be properly adoptable. As this would typically also involve inter-domain relationships due to dependencies, the models need to be designed to support such cross-domain handling.

Designing, managing, providing, and using Understandable, Findable, and Updatable models are considered important tasks of a data space environment and its infrastructure. The fourth scenario of Operate then strongly addresses the control and operation of the involved systems by means of such models, highlighting the needs for standardization of control functions and semantic protocols, which need to be linked to respective functional information models.

3 IDSA Approach to Semantic Interoperability in Data Spaces

3.1 Main contributions of the IDS-RAM

The IDS-RAM[2] provides a comprehensive view of the structure and the concepts in a Data Space. Using a layered approach, the different concepts are described.

The ecosystem of the IDS comprises several basic tasks being carried out by the various participants as descripted in the IDS-Reference-Architecture Model [2]. The set of these tasks can be derived from relevant objects in the IDS and the activities along the respective life cycle. Among those objects are the Vocabularies, which are ontologies, reference data models, or metadata elements that can be used to annotate and describe datasets, usage policies, apps, services data sources etc.

The Vocabulary Intermediary technically manages and offers vocabularies (i.e. ontologies, reference data models, or metadata elements). The Vocabulary Intermediary typically assumes the basis roles of the Vocabulary Publisher and Vocabulary Provider. Vocabularies are owned and governed by the according Standardization Organization.

Vocabularies can be used to annotate and describe data assets. These data assets may comprise at least:

- Information Model of the International Data Spaces, which is the basis for the description of data sources[12].

- Domain-specific vocabularies are essential for the scalability and success of the IDS. Domains are e.g. represented in the quite common set of linked open data.

- Legal terms: To describe usage policies and to enable smart contracting, legal terms must be coded in a machine-readable and -understandable manner. The IDS Information Model defines the Open Digital Rights Language (ODRL) to describe usage policies. Still, IDS communities such as a (closed) supply chain network or a domain-specific IDS initiative could define additional (complementary or alternative) vocabularies, e.g. depict the International Commercial Terms (Incoterms) as an ontology.

There is no dedicated or exclusive role that creates vocabularies. Usually, standardization organizations such as ISO/IEC, CEN/CENELEC, IEEE etc., but also industrial associations define standards that can be formulated as a vocabulary (Vocabulary Creators and Owners). Except the IDS information model, there can be multiple vocabularies describing the same context (e.g. different types of smart contracts or usage policy descriptions). A single vocabulary for the same context supports standardization and, thus, compatibility efforts. Multiple vocabularies provide flexibility and competitiveness.

In specific IDS-based ecosystems, domain-specific adaptations – also known as Application Profiles – of the Information Model may be used to describe Resources, Participants, infrastructure, and other constituents of an International Data Space.

Further, independent domain-specific Vocabularies, which are not necessarily derived from the IDS Information Model, may be used to describe the Content of a Resource and the Concepts addressed by a Resource, as detailed in the respective sections below.

The Vocabulary Hub in IDS addresses, as described above, the need for managing vocabularies during the lifecycle. From the perspective of a data provider and a data consumer, two phases should be distinguished, the Design Phase and the Runtime Phase.

During the creation of the Data Offering the Data Provider may reuse, as described above, existing standards for the (semantic) description of the data itself or create a (semantic) description of the data. These Vocabularies can be published to a Vocabulary Hub and linked to the self-description. This Design-time step supports the semantic interoperability in Data Spaces. While semantic models for the description of data in data spaces are in general a good practice, Vocabularies can also make use of other concepts.

The IDS Metadata Broker does not serve Vocabularies but provides a reference to a vocabulary and, if required, a reference to a Vocabulary Hub, included in the Self-Description during Runtime when a connector is searching for a data provider or a data set. The Data Consumers connector may verify if the data is provided by using a vocabulary that is consumable by the connector, when querying an IDS Metadata Broker or when querying the Self-Description directly from a Data Providers Connector. If the data is not provided in a consumable way, the connector may:

- request the data in a different format from the data provider or search and invoke another service that can conduct a transformation of the data according to another data scheme,
- implement the required structures (interfaces) to consume the data. As this could be a manual task to implement the required interfaces or code fragments, this could be a time-consuming task,
- or choose a different data provider, which provides the required data in a schema and format that is usable by the data consumer.

When a Vocabulary related to the data is provided by Data Provider, the Data Consumer may validate the provided schema by reasonable means before initiating the contract negotiation.

The steps described above are summarized in Figure 3below, distinguishing the two phases of Design Time of data assets and runtime of data exchange.

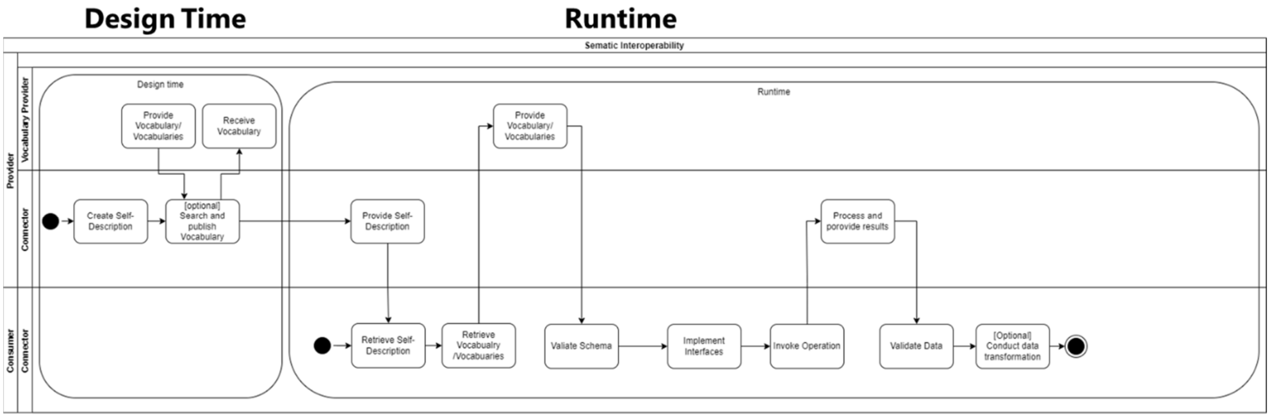

Figure 3 Activities in semantic interoperability in Data Spaces (source: IDS-RAM 4

The detailed description of the technical processes are part of the IDS-RAM[2] section 3.4.

3.2 Duties of Individual Data Space instances to achieve semantic interoperability

The IDSA Paper on Inter and Intra Data Space Governance [13] discusses duties of data space instances, i.e., conducted via the Data Space Authority as described in the IDSA Rulebook [5], and the governance aspects of data spaces in general to achieve interoperability between the instances. Semantic interoperability is one aspect of intra and inter data space governance. Two aspects of semantic interoperability are relevant for the data space instances, the management of common semantic data models beyond the generic enablement of the IDS information model [12] and semantic management data services.

3.2.2 Common semantic data models

The IDS Information Model based on DCAT and ODRL is the basic semantic model for IDS-based data spaces. 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.

The Service Providers implement the domain specific model in the related Data App (as part of the IDS Connector), and this implementation may be certified by the Data Space Instance to ensure interoperability and compliance to the rules and policies of this Data Space instance. Finally, the Data Space Instance is responsible for support on the domain-specific data models. Those may be managed with different governance models.

3.2.3 Semantic management data services

The Semantic Management Data services are used within data space instances to map different semantic models to the agreed upon common semantic data model (in some context, e.g. cyber security, mapping might be not the right methodology, as this might lead to unclarity). To integrate End-users, a semantic mapping between the End-users backend systems and the common semantic data model is required. This mapping may be standardized and certified by the Data Space Instance and is developed by the Service Providers. The Service Providers facilitate the connection of the End-users and therefore are also the main support contact point. A Vocabulary Hub may support the semantic management data services with publication and provisioning of semantic models. The Data Space instance may provide one or multiple vocabulary hubs or reference to existing instances.

4 Selection of best practices and good examples

Following the general introduction of the need for semantic interoperability and the approach for Data Spaces in general, this section will provide illustration of such scenarios and insights on the current developments by making use of real-life use cases. The selected use cases span through different domains.

4.1 Use Case: Port Logistics – On the need for interoperability

Port ecosystems include several stakeholders competing fiercely in an environment with a robust data protection policy. However, at the same time, these stakeholders have complementary interests. These characteristics lead to operational inefficiencies due to their reluctance to share data for the common good.

In the context of a European project that aims towards Smart Green Logistics for the future of ports, International Data Spaces (IDS) can be helpful in port logistics, becoming an enabler for the digital transition of ports, providing a common language and framework for organizing and sharing information across different systems and stakeholders [20].

In the context of ports, which encompass a multifaceted array of systems and stakeholders, including shipping companies, logistics providers, customs agencies, and port authorities, each with unique information management systems and formats, achieving interoperability can be arduous.

However, a Vocabulary Hub can provide a standardized way to represent and disseminate information across these different systems, thereby fostering interoperability and mitigating any friction in the exchange of information between the stakeholders. Automation has become vital as ports optimize their operations and minimize costs.

Given the complex information environment, data quality can pose a significant challenge in the port ecosystem. Nevertheless, a Vocabulary Hub can help to address this challenge by providing a set of standardized terms and concepts, reducing ambiguity, and ensuring consistency across different systems and data sources.

The sheer volume of data ports generate necessitates an organized framework to make sense of it all. A Vocabulary Hub can provide this structure, enabling more effective analysis and decision-making based on the insights gleaned from the data, including optimized energy grid management.

4.2 Use Case: Clothing Industry

The textile and clothing industry has different stakeholders with different roles and responsibilities along the value chain that requires specific and critical information to perform their tasks.

First, there are the fashion brands that design products. Then critical information about the product, digital product, containing information about the product image and design, technical draws, color and patterns, sizes, and raw material quality requirements, among other information is shared with manufacturing companies that are responsible for the production of the products. The manufacturing company defines the detailed product specifications and process flow diagrams. Here, the raw material suppliers as well as the different manufacturing sub-contractors are defined, the production plan is defined according to the manufacturing capacities, expertise, and availability. Normally these textile Manufacturing companies need to manage a network with more than 30 industrial companies, selecting the main experts in different operations, according to the production requirements of each specific product.

As previously described, the clothing industry is truly dependent on effective value chain management, from the fashion company to the subcontracted factory that will execute each of the production stage. Today the data exchange between the fashion company and the manufacturing company, as well as the horizontal integration between the manufacturing company and the sub-contractors are done in an ad-hoc way, using email or excel files to plan and monitor the production process.

In the future, the vision is to increase flexibility in the clothing industry, reduce dependence from centralized production and explore the manufacturing capacity as a service concept to have distributed production value chain, exploring local networks, and thus reduce the need to have global logistics.

This reality is not aligned with the problems and inefficiencies previously described. It is necessary to explore new, trust and effective communication channels that enable different value chain stakeholders to have access and share sensible information about product and production processes and costs.

In this sense, to setup local value chain capable to deliver the needs from a specific region, there is a critical need to streamline the data and information communication between the different actors, towards the value chain creation, management, and optimization. This data must be exchanged according to a unique vocabulary and semantics that is understandable by all actors. Even new actors to the network, that will be identified and invited to the network, should be able to easily define contract that will give them the legal rights to access to information from the fashion companies, but also be able to understand this data in an unequivocally way.

4.3 DATES. European Data Space for Tourism

DATES [14] is a EU project that aims to explore approaches and options for the deployment of a secure and trusted tourism data space, ensuring transparent control of data access, use and re-use. The project focuses on the development of governance and business models, while providing a shared roadmap that will ensure the coordination of the tourism ecosystem stakeholders and the connection between data ecosystems at EU level and interconnected data spaces in other sectors.

The project promotes the vision of a prosperous tourism data space and recommends clear strategies on how to inspire and motivate all key tourism stakeholders to collaboratively build a powerful interconnected tourism data space. Besides providing leadership and practical advice about how every stakeholder in the tourism value chain can contribute and utilize data streams, the added benefits of a European Tourism Data Space will be highlighted from each stakeholder's perspective. In more general terms, DATES supports the digital transformation of the sector, fostering competitiveness, resilience, and sustainability as key success factors to maintain Europe's leading role.

4.4 DEMETER

The DEMETER AIM (Agriculture Information Model) which is openly accessible through Open Geospatial Consortium (OGC) [15] and the LIRMM AgroPortal repository [16], boasts various benefits and features, with the most significant being its ability to achieve interoperability with established ontologies and systems. Alongside the mappings described earlier in the cross-domain AIM ontology, AIM also includes semantic interoperability support and mappings to other frameworks such as FIWARE, SAREF4AGRI, ADAPT, INSPIRE/FOODIE, AGROVOC, and the EPPO2 modelling approach. The AIM architecture facilitates scalability through its modular design, which allows for updates and new additions to the model in a straightforward manner. The model aims to ensure interoperability between various data types including: geospatial, environmental, water, soil, crop, animal welfare, product, machinery and farm/logistics.

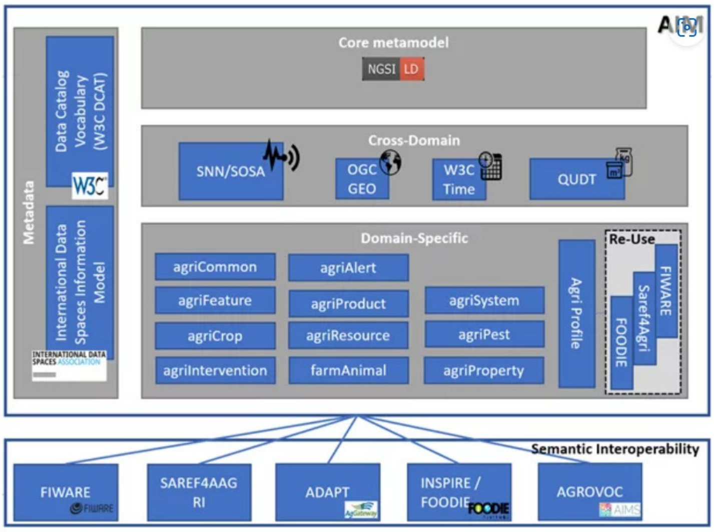

Figure 4 DEMETER Agricultural Information Model [15]

More information on the common interoperability mechanisms used in the project are accessible via the project's website [17].[18]. In this document [19], one can also find more specific descriptions about DEMETER AIM, as well as its position towards IDS Information Model [12] and other examples of semantic interoperability mechanisms from agrifood sector (such as ISOBUS and AGROXML). The model is also on its way to become an OGC standard. [19], one can also find more specific descriptions about DEMETER AIM, as well as its position towards IDS Information Model and other examples of semantic interoperability mechanisms from agrifood sector (such as ISOBUS and AGROXML). The model is also on its way to becoming an OGC standard.

DEMETER AIM will also play a key role in DIVINE project [17], [20] (Demonstrating Value of Agri Data Sharing for Boosting Data Economy in Agriculture) that started in October 2022. This project will implement an agriculture dataspace ecosystem by following IDS RAM as the backbone of its architecture, ensuring secure and sovereign data sharing (data exchange via IDS Connectors) and DEMETER AIM will be the core mechanism to be used to ensure interoperability within agriculture stakeholders. while ensuring interoperability via the use of DEMETER AIM. One of the pillars of the project that is under development is called ADAM (DIVINE Agriculture Data Model). This will be positioned as a semantic data model to be used by all data providers, consumers, and pilots within the project's ecosystem. It will incorporate and extend existing agrifood ontologies, vocabularies, and information models (e.g., SAREF4AGRI, NGSI-LD agri data model of FIWARE, IDS Information Model, ADAPT, INSPIRE, AGROVOC, AgroXML/AgroRDF, AIM), as well as for related domains such as agri-machinery, weather data, IoT, supply chain, data catalogs, user generated data, etc.

In construction project environments, permit management processes, including compliance checks, are very complex.

These processes implicate data sharing between many actors in the value chain. Architects and engineers that are involved in the design idea, construction companies, subcontractors or suppliers that are involved in the construction itself and certification companies and public authorities that are part of the expedition and approval of permits.

4.5 Digichecks

In turn, while there is no common standard within different regions of authorities related to the format of the regulations, there is isolated data between different stakeholders and is usually stored in silos, which delivers information in specific formats that in some cases cannot be processed by machines. Examples of these are rvt, dwg and pdf formats. Therefore, processing this information in permit management processes is not automatic and in some case requires manual checks that lead to excessively long waiting times.

To automate these processes, the DigiChecks project (https://digichecks.eu/) aims to develop a solution to provide flexibility, ease-of-use and efficiency to the permit validation and approval system in the construction project environments. This solution addresses the sharing of data related to the management of the construction permits regardless of the country, region or municipality, following European values and standards of federation, openness, interoperability and digital sovereignty. In this regard, a critical aspect is the semantic interoperability. That is, the definition of a shared language for permitting that formalized in a permit ontology enables to map data from various sources into a common structure and make it automatically processable by a machine in a repeatable manner.

4.6 Data Spaces Support Centre (DSSC)

The DSSC is a project funded by the European Commissions that operationalizes the European strategy for data and facilitates common data spaces that together create an interoperable data sharing environment. The aim is to enable data reuse and secondary use within and across sectors and thus contributing to the European economy and society.

The DSSC explores the needs of the data spaces initiatives, the common requirements, and best practices. By distilling existing solutions, integrating what works, and closing gaps of what is missing, the DSSC will deliver a data spaces blueprint. This is composed of common building blocks which encompass the business, legal, operational, technical, and societal aspects of data spaces. The blueprint continuously evolves with a user-centric approach, as the result of co-creation with the stakeholders. As this paper depicts, semantic interoperability is a major concern for data space and will be reflected in the work of the DSSC and the blueprint process.

4.7 The European Landscape of semantic interoperability

The EU-level agenda for promoting sematic (and other) interoperability is being advances particularly by DG DIGIT and its Interoperable Europe initiative (formerly known as the ISA2 programme), which includes the SEMIC (Semantic Interoperability Community) action.

SEMIC „develops solutions to help European public administrations perform seamless and meaningful cross-border and cross-domain data exchanges. The provision of digital cross-border public services requires the exchange of data between public administrations of different EU countries. Semantic interoperability is a fundamental enabler of such exchanges. It is crucial to agree on the use of common semantic standards, promote transparent and well-documented metadata policies and increase the visibility and reuse of existing semantic interoperability solutions.“ [21]

The interoperability unit of DG DIGIT (B.2), in addition to coordinating the SEMIC action, is the responsible unit for the European Commission proposal for the Interoperable Europe Act (IEA, published in November of 2022). The IEA aims to strengthen cross-border interoperability and cooperation in the public sector across Europe. In particular, it aims to tackle three limitations of the current landscape:

- Inefficient governance of interoperability efforts between different policies on the different administrative levels of the EU and its member states.
- Lack of common minimum interoperability specifications, shared solutions, and open standards.
- Lack of an 'interoperability-by-default' approach in the design and implementation of legislation and policies on various levels and in different contexts.

5 Outlook

In summary, this paper has delved into the crucial concept of semantic interoperability within the realm of Data Spaces, aligning it with the European Interoperability Framework and the facets outlined in ISO/IEC 21823-1:2019.

The IDS Reference Architecture Model 4 has played a pivotal role in elucidating the essential activities required for achieving semantic interoperability, presenting a clear path, and understanding of the involved components. This framework significantly contributes to the successful implementation of interoperable data spaces.

In the concluding chapter, we offer a valuable overview of best practices and real-world use cases, providing practical insights into the application of semantic interoperability. As organizations grapple with the intricacies of data integration, the insights shared in this paper serve as a guide for fostering effective collaboration and interoperability in the dynamic landscape of data spaces."

As we look ahead, the exploration of semantic interoperability in Data Spaces continues to be a dynamic and evolving endeavor. Building upon the insights provided in this paper, our journey extends into the future with a focus on fostering collaboration, innovation, and standardization within the domain. Notably, this paper aligns with and supports the broader vision and mission of the expert workshop series conducted in 2022 and 2023.

These workshops, documented and detailed on the webpage semantic.internationaldataspaces.org, serve as critical forums where researchers, decision makers, and practitioners converge to deliberate on the development and operation of European (common) Data Spaces, data markets, and other web-based data management systems. The primary objectives encompass discussing, refining, and specifying requirements for both syntactic and semantic interoperability. The outcomes of these discussions are committed to being freely available under open licenses and open access.

Furthermore, the workshops aim to contribute to the landscape of existing tools and vocabularies by identifying and cataloging concrete tools relevant to the field. Again, these outcomes are intended to be shared openly, supporting a collaborative and transparent approach to advancing interoperability in data spaces.

A key focus moving forward will be on addressing the need for related standardization, recognizing its pivotal role in establishing a robust foundation for interoperability. This paper, aligned with the workshop series, seeks to contribute substantively to these ongoing efforts, fostering a community-driven approach to advancing semantic interoperability in the realm of data spaces. As we collectively navigate the complexities of data sharing, trading, and collaboration, these initiatives aim to guide and shape the future of interoperable data ecosystems, ensuring accessibility, transparency, and efficiency for all stakeholders involved.

References

[1] MyData Global, “Data Sovereignty.” [Online]. Available: https://www.mydata.org/2022/09/26/data-sovereignty/
[2] International Data Space Association, “IDS-Reference Architecture Model 4.” [Online]. Available: https://docs.internationaldataspaces.org/ids-ram-4/
[3] European Commission. Directorate General for Informatics., *New European interoperability framework: promoting seamless services and data flows for European public administrations.* LU: Publications Office, 2017. Accessed: Mar. 27, 2023. [Online]. Available: https://data.europa.eu/doi/10.2799/78681
[4] “Data Connector Report,” International Data Spaces. Accessed: Mar. 27, 2023. [Online]. Available: https://internationaldataspaces.org/data-connector-report/
[5] “IDSA Rulebook.” [Online]. Available: https://docs.internationaldataspaces.org/ids-knowledgebase/v/idsa-rulebook/idsa-rulebook
[6] “ISO/IEC 21823-1:Internet of Things (IoT) — Interoperability for Internet of Things Systems — Part 1: Framework.”
[7] “ISO/IEC CD 21823-2: Internet of Things (IoT) — Interoperability for Internet of Things Systems — Part 2: Transport Interoperability.”
[8] “ISO/IEC CD 21823-3:Internet of Things (IoT) — Interoperability for Internet of Things Systems — Part 3: Semantic Interoperability.”
[9] Martin Bauer et al., “Towards semantic interoperability standards based on ontologies. AIOTI white paper.,” Jun. 2019.
[10] DIN/DKE, “Artificial Intelligence Standardization Roadmap, 2nd edition,” Dec. 2022. [Online]. Available: https://www.dke.de/standardization-roadmap-ai
[11] IEC, “IEC White Paper: Semantic interoperability: challenges in the digital transformation age.” Accessed: Mar. 19, 2021. [Online]. Available: https://www.iec.ch/basecamp/semantic-interoperability-challenges-digital-transformation-age
[12] International Data Space Association, “IDS Information Model.” [Online]. Available: https://w3id.org/idsa/core
[13] International Data Space Association, “Governance for Data Space Instances – Aspects and Roles for the IDS Stakeholders,” International Data Spaces. Accessed: Mar. 27, 2023. [Online]. Available: https://internationaldataspaces.org/publications/position-papers/
[14] “DATES. European Data Space for Tourism.” [Online]. Available: https://www.tourismdataspace-csa.eu/
[15] “OGC-DEMETER.” [Online]. Available: https://www.ogc.org/initiatives/demeter/
[16] “DEMETER Ontology.” [Online]. Available: https://agroportal.lirmm.fr/ontologies/DEMETER-AIM
[17] “an update on the agriculture information model.” [Online]. Available: https://h2020-demeter.eu/an-update-on-the-agriculture-information-model/
[18] DEMETER Project, “D2.1 Common Data Models and Semantic Interoperability Mechanisms.” [Online]. Available: https://h2020-demeter.eu/wp-content/uploads/2020/10/DEMETER_D21_final.pdf
[19] R. Palma *et al.*, “Agricultural Information Model,” in *Information and Communication Technologies for Agriculture—Theme III: Decision*, vol. 184, D. D. Bochtis, C. G. Sørensen, S. Fountas, V. Moysiadis, and P. M. Pardalos, Eds., in Springer Optimization and Its Applications, vol. 184. , Cham: Springer International Publishing, 2022, pp. 3–36. doi: 10.1007/978-3-030-84152-2_1.
[20] “DIVINE Project.” [Online]. Available: http://divine-project.eu/home
[21] “SEMIC.” [Online]. Available: https://joinup.ec.europa.eu/collection/semic-support-centre/about