



Position Paper | Version 1.1 | November 2025

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-Straße 80
44227 Dortmund
Germany

Copyright

International Data Spaces Association,
Dortmund 2025



<https://creativecommons.org/licenses/by/4.0>

Editor

Sebastian Steinbuss, IDSA

Cite as

Steinbuss, S. et al., Semantic Interoperability
in Data Spaces, International Data Spaces
Association, 2025

<https://doi.org/10.5281/zenodo.17630664>

Authors & contributors

António Henrique Almeida, INESC TEC
Wouter van den Berg, TNO
Sthefan Gabriel Berwanger, INESC TEC
Sonia Bilbao Arechabala, Tecnalia
Izaskun Fernandez, Tekniker
Gonzalo Gil Inchaurrea, Tekniker
Pieter Klinker, TNO
Viivi Lähteenoja, 1001 Lakes
Andreas Müller, Schaeffler
Michiel Stornebrink, TNO
Sebastian Steinbuss, IDSA
Anil Turkmayali, IDSA
Petteri Kivimäki, NIIS
Philippe Calvez, Acatech
Peter Koen, Microsoft

Contents

1. Introduction	4
1.1 Types of interoperability	4
1.2 On the need for semantic interoperability in data spaces.....	6
2. General approach to semantic interoperability.....	9
3. IDSA approach to semantic interoperability in data spaces.....	12
3.1 Main contributions of the IDS-RAM	12
3.2 Abilities of the Dataspace Protocol.....	14
3.3 Data Catalog Vocabulary (DCAT) – version 3.....	15
3.4 Open Digital Rights Language (ODRL).....	18
3.5 Duties of individual data space instances to achieve semantic interoperability	19
3.5.1 Common semantic data models	20
3.5.2 Semantic management data services	20
4. Selection of best practices and good examples	21
5. Outlook	22
6. References	24

1. Introduction

Data sovereignty constitutes a core aspect of the International Data Spaces Association (IDSA). 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]. To this end, the International Data Spaces Association proposes a Reference Architecture Model for this capability and related aspects, including requirements for secure and trusted data sharing in business ecosystems.

The IDSA Rulebook provides the principles for trusted data sharing using data spaces, including a clear guideline for the mandatory and optional requirements of data spaces. Among these, vocabularies and semantic models are identified as part of the mandatory foundation of data spaces (see Figure 1).

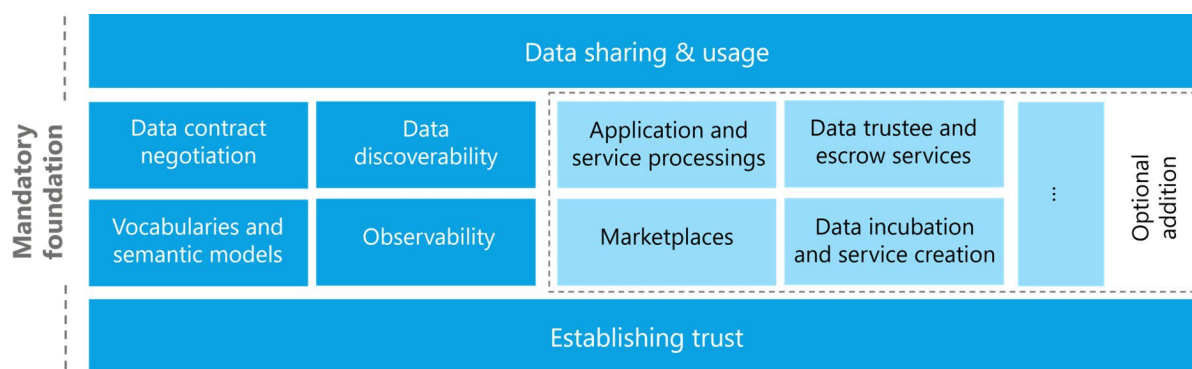


Figure 1 Foundational concepts of Data Spaces (source: IDSA Rulebook)

This paper focuses on the semantic interoperability aspect of data spaces, which constitutes a key requirement in data spaces.

1.1 Types of interoperability

Interoperability can be addressed on distinct levels. The New European Interoperability Framework (EIF) [3] defines an interoperability model with four layers of interoperability: technical, semantic, organizational, and legal.

The layers of technical and semantic interoperability are covered by data connectors, which are defined in the IDS-RAM [2] and are further described in the IDSA Data Connector Report [4]. As detailed in the remainder of the document, both may require interaction with additional services or components. Legal and operational interoperability, however, can be achieved by the policies and rules of a specific data space instance and are typically managed by a data space governance authority. More information on this can be found in the IDSA Rulebook[5].

In addition to the EIF, which is applicable to all digital public services, ISO/IEC 21823-1:2019 Internet of things (IoT) — Interoperability for IoT systems [6] and ISO/IEC 19941:2017 Information technology — Cloud computing — Interoperability and portability introduce a five-facet model for the interoperability of IoT systems and for interoperability in cloud



computing and distributed systems, addressing transport, syntactic, semantic, behavioral and policy interoperability. Although they use slightly different names, the frameworks address similar concepts.

Technical interoperability in the EIF covers both transport and syntactic interoperability of ISO/IEC 21823-1 and ISO/IEC 19941. On the one hand, transport interoperability [7] is responsible for 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. Hence, aspects of technical interoperability include interface specifications, interconnection services, data integration services, data presentation and exchange, and secure communication protocols.

Based on this, semantic interoperability enables the exchange of data between entities using mutually understood data or information models, i.e., their semantics [8]. According to [3], [9], semantic interoperability is reached when interacting systems attribute the same meaning to an exchanged data entity, ensuring consistent handling of the data across systems regardless of individual data formats. This consistency of meanings can be achieved using existing standards or agreements on the format and meaning of data, or it can be derived dynamically using shared vocabularies either in a schema form and/or in an ontology-driven approach.

So, in short, while transport interoperability deals with the delivery of the data (i.e., data transmission) and syntactic interoperability enables reading of the data according to a known format and grammar, semantic interoperability is concerned with the meanings of the data, enabling its unambiguous interpretation and understanding.

On top of this, organizational interoperability in the EIF 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. On 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 and in ISO/IEC 1994 ensure that the actual result that is exchanged leads to the expected outcome.

Finally, legal (EIF) or policy (ISO/IEC 21823-1:2019/ISO/IEC 19941) interoperability is about ensuring the successful collaboration of and between organizations that operate under different legal frameworks, policies, and strategies.

Table 1 provides a summary of the EIF and ISO/IEC 21823-1:2019/ISO/IEC 1991.



Table 1 Comparison EIF and ISO/IEC 21823-1:2019/ISO/IEC 19941:2017

European Interoperability Framework	ISO/IEC 21823-1:2019 ISO/IEC 19941:2017	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 and interrelations
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

With regard to the most basic task of data spaces – the controlled exchange of well-understood data between participants – this document provides an overview of means for and examples of specifically semantic interoperability in data spaces.

1.2 On the need for 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. Unclear meaning of shared data can lead to miscommunication and misinterpretation, resulting in errors and poor decisions. Hence, organizations that must deal with ambiguous data need to spend a lot of effort in mapping that data to the formats and structures expected by their IT systems. Being time-consuming and costly, the lack of shared meaning of data forms a major barrier to data sharing and, therefore, to the realization of the Digital Single Market strategy¹. This

¹ https://ec.europa.eu/growth/tools-databases/enorm/mandate/614_en



need is also clearly addressed in *ISO/IEC 20151 Information technology — Cloud computing and distributed platforms — Dataspace concepts and characteristics*².

The ability of IT systems to exchange data with unambiguous, shared meaning is called semantic interoperability. As an essential requirement for federated data networks, such as IDS and data spaces in general, it requires data providers and consumers across the network to describe their data offerings or needs using explicit references to a common vocabulary. The commonality is important to enable understanding of data, the receiving party is unfamiliar with. In such cases, the need to spend integration efforts on their end remains controllable. Similarly, if relational semantics are lost or inadequately expressed during exchange, receiving systems may misinterpret data context, leading to incorrect conclusions despite syntactically accurate transfer. Vocabularies become common (shared) through the process of standardization (using different means). This standardization must address not only data element definitions but also the semantic models that capture how these elements relate to one another within their domain context.

Linked Data is a method of publishing structured data so that it can be interlinked and become more useful through semantic queries, using standards like RDF. It enables data from different sources to be connected and queried in data spaces. This concept can leverage the abilities of interoperable data spaces.

Many industries and 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 that are freely accessible and implementable. 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 that there is often a limit to the level of semantic interoperability that can be achieved. Every member in a business ecosystem comes 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 continue to do 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, however, achieving semantic interoperability requires the use of semantic technologies to eventually create Linked Data. Semantic technologies such

² <https://www.iso.org/standard/86589.html>



as RDF, SHACL, OWL and SKOS allow for enriching data with meaning by creating links to other datasets and vocabularies, ultimately 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 [2, section 3].

The next section sheds some light on the steps necessary to achieve this goal.



2. General approach to semantic interoperability

The rise of data spaces goes hand in hand with the demand for connecting, interpreting, and processing large volumes 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 its meanings. 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 a 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. This understanding must encompass not only individual data elements, but also critically, the semantic relationships between them. Without capturing these relational semantics, systems cannot perform sophisticated reasoning or maintain data integrity across different operational contexts.
- **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. Advanced finding capabilities leverage semantic relationships to enable transitive queries, path-based discovery, and contextual search that goes beyond simple attribute matching. For instance, finding all data "related to" a specific entity requires understanding the semantic graph of relationships, not just direct associations.
- **Updating** found data and models based on given criteria to keep track of various changes influencing the entities and data. Updates must preserve relational integrity and propagate semantic implications across connected data elements. When one entity changes, systems must understand which related entities are affected and how relationship semantics influence the cascading updates required to maintain consistency.
- **Operating** on the updatable entities and data based on given criteria, i.e., performing operations of different complexities. Operations benefit from semantic relationship understanding to enable advanced analytical capabilities such as causal chain analysis, constraint propagation, predictive inference, and cross-domain pattern recognition. These sophisticated operations are only possible when both data semantics and their relational context are preserved and computationally accessible.



Mastering these steps allows for realizing many data-driven use cases that nowadays need considerable manual efforts for system integration to be 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 the 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 leading to interoperability can still prove beneficial in many ways. On a purely technical level, this leverages a clearer understanding of the data and their relationships, 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 them 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 meaning that the development, operation and governance processes involved 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 a 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 a 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 need 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 described in the IDS-Reference-Architecture Model [2]. These tasks are derived from key objects within the IDS and their respective lifecycle activities. Among those objects are the Vocabularies, comprising ontologies, reference data models, and metadata elements that serve to annotate and describe diverse components within the system. These Vocabularies enable semantic interoperability across different knowledge domains by providing standardized ways to characterize datasets, define usage policies, describe applications and services, and document data sources

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 respective 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.
- IDS Information Model:
- The foundational vocabulary that provides standardized descriptions for data sources across the entire IDS ecosystem. This model serves as the common semantic framework enabling interoperability between all IDS participants.
- 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.
- Domain-Specific Vocabularies
- Essential for IDS scalability, these vocabularies capture the specialized knowledge and terminology of particular sectors or industries. They are often represented through established linked open data standards, allowing domains to maintain their specific semantic requirements while ensuring compatibility with the broader IDS framework.
- 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
- Legal and Contractual Vocabularies:



- These vocabularies encode legal terms in machine-readable formats to enable automated usage policy enforcement and smart contracting. While the IDS Information Model adopts the Open Digital Rights Language (ODRL) as its standard for usage policies, specific IDS communities—such as closed supply chain networks or domain-specific initiatives—may develop complementary vocabularies. For example, International Commercial Terms (Incoterms) can be represented as an ontology to support automated trade agreements within supply chain contexts. A data space 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 vocabulary. 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). Apart from 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. Mappings of similar terms and classes between different vocabularies and ontologies enable interoperability among those different understandings or interpretations of the same context or domain.

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 (new) (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 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 Provider's 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 2 below, distinguishing the two phases of Design Time of data assets and runtime of data exchange.

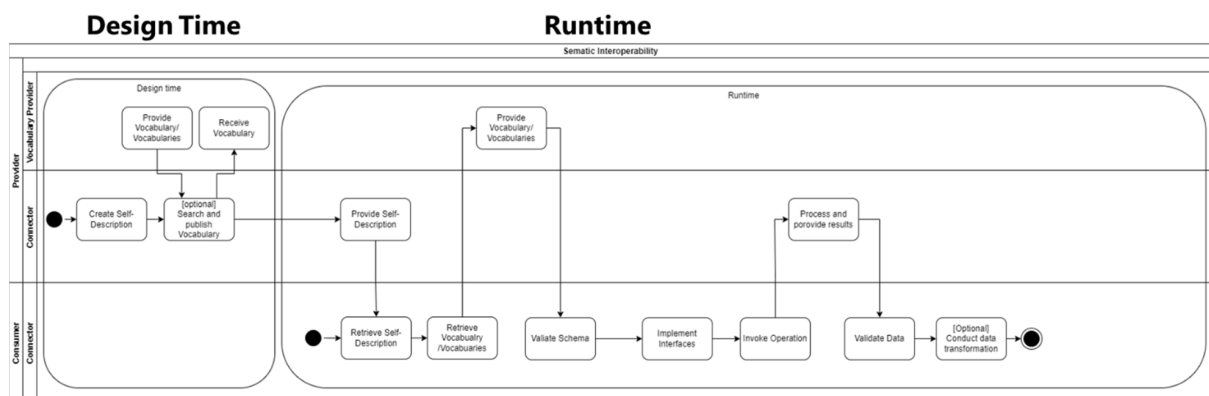


Figure 2 Activities in semantic interoperability in Data Spaces (source: IDS-RAM 4)

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

3.2 Abilities of the Dataspace Protocol

The Dataspace Protocol is the key element for interoperability in data spaces with a strong focus on transport and syntactical interoperability by the definition of message sequences and their structure. The fundamental behavior of data providers and data consumers is expressed by the definition of state machines. However, semantic interoperability is enabled by the Dataspace Protocol also to a certain extent by the usage of certain semantic models. The Dataspace Protocol does not describe mechanisms for the semantics of the data itself, but provides a foundation for metadata aspects, data catalogs, and policies.

The **Dataspace Protocol** is used in the context of data spaces³ as described and defined in the specification document⁴ with the purpose to support *interoperability*. In this context, the

³ <https://eclipse-dataspace-protocol-base.github.io/DataspaceProtocol/2025-1-RC4/#dfn-dataspace>

⁴ <https://eclipse-dataspace-protocol-base.github.io/DataspaceProtocol/>



specification provides fundamental technical interoperability for participants⁵ in data spaces.

This specification builds on protocols located in the ISO OSI model (ISO/IEC 7498-1:1994)⁶ layers, like the Hypertext Transfer Protocol (HTTP). The purpose of this specification is to define interactions between systems independent of such protocols, but describing how to implement it in an unambiguous and extensible way. To do so, the messages⁷ that are exchanged during the process are described in this specification and the states and their transitions are specified as state machines, based on the key terms and concepts of a data space. On this foundation, the bindings to Data Transfer Protocols⁸, like Hypertext Transfer Protocol Secure (HTTPS), are described.

Note: This specification does not cover the data transfer as such. While this is controlled by the Transfer Process Protocol, 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.

The classes and definitions used in this specification are reused from different standards and specifications as much as possible, in particular, Data Catalog Vocabulary (DCAT) [*vocab-dcat-3*]⁹ and Open Digital Rights Language (ODRL) [*odrl-model*]¹⁰. As, however, the external definitions allow different interpretations or provide more attributes than required, this specification is leveraging profiles¹¹ of the original definitions rather than the complete original expressiveness. However, not every standard-compliant class might be compliant to the data space profile. They are not separate artifacts but implicitly contained in the JavaScript Object Notation (JSON) schemas for the message types¹² of this specification.

3.3 Data Catalog Vocabulary (DCAT) – version 3

DCAT is an RDF-based vocabulary developed by the W3C to describe datasets and data services in data catalogs on the Web. Its primary goal is to enhance metadata interoperability, enabling federated search and aggregation across distributed catalogs. The original DCAT was standardized in 2014, followed by DCAT v2 in 2020, and now DCAT v3, published as a W3C Recommendation in August 2024¹³.

DCAT v3 builds upon community feedback and evolving use cases, introducing enhancements while maintaining backward compatibility with DCAT v2. It extends the vocabulary to support dataset versioning, dataset series, and checksum validation, among

⁵ <https://eclipse-dataspace-protocol-base.github.io/DataspaceProtocol/2025-1-RC4/#dfn-participant>

⁶ <https://www.iso.org/standard/20269.html>

⁷ <https://eclipse-dataspace-protocol-base.github.io/DataspaceProtocol/2025-1-RC4/#dfn-message>

⁸ <https://eclipse-dataspace-protocol-base.github.io/DataspaceProtocol/2025-1-RC4/#dfn-data-transfer-protocol>

⁹ <https://eclipse-dataspace-protocol-base.github.io/DataspaceProtocol/2025-1-RC4/#bib-vocab-dcat-3>

¹⁰ <https://eclipse-dataspace-protocol-base.github.io/DataspaceProtocol/2025-1-RC4/#bib-odrl-model>

¹¹ <https://eclipse-dataspace-protocol-base.github.io/DataspaceProtocol/2025-1-RC4/#dfn-profile>

¹² <https://eclipse-dataspace-protocol-base.github.io/DataspaceProtocol/2025-1-RC4/#dfn-message-type>

¹³ <https://www.w3.org/TR/vocab-dcat-3/>



other features. The vocabulary integrates terms from established ontologies (e.g., FOAF, DCTerms, SPDX) and is designed to be modular and extensible.

DCAT v3 introduces several functional improvements:

- Versioning Support: New properties like `dcat:version`, `dcat:previousVersion`, and `dcat:hasCurrentVersion` enable precise tracking of dataset evolution.
- Dataset Series: The `dcat:DatasetSeries` class allows grouping datasets that form a logical sequence.
- Checksum Validation: Integration of SPDX terms (`spdx:Checksum`) supports integrity verification of distributions.
- Inverse Properties: Enhances bidirectional linking between resources.
- Relaxed Constraints: Facilitates broader adoption by allowing more flexible metadata modeling.
- Interoperability Enhancements: Continued alignment with external vocabularies and support for decentralized publishing and federated search.
- These features collectively improve metadata expressiveness, discoverability, and integration across heterogeneous data ecosystems.

DCAT v3 has seen growing adoption across the public sector and research domains. The DCAT-US v3 profile, developed under the FAIRness Project, aligns federal metadata practices with DCAT v3, supporting documentation of business, statistical, and geospatial data. Similarly, the DCAT-AP v3 profile in Europe facilitates cross-border metadata harmonization, with mature implementations in domains like health, mobility, and smart cities.

Vendor support and tooling (e.g., SHACL validation, JSON-LD serialization) are expanding, and community engagement through GitHub and webinars fosters collaborative evolution. The standard is increasingly seen as a cornerstone for metadata interoperability in data spaces and open data portals.

DCAT v3 plays a pivotal role in enabling semantic interoperability by providing a shared, machine-readable vocabulary for describing datasets. Through its RDF foundation and alignment with Linked Data principles, DCAT ensures that metadata can be understood, linked, and reused across systems and domains.

Application profiles like DCAT-AP extend DCAT with domain-specific constraints and controlled vocabularies, supporting cross-domain metadata harmonization. This facilitates the creation of interoperable data spaces, where metadata from diverse sources can be aggregated, validated, and queried uniformly.

DCAT's semantic rigor, extensibility, and alignment with FAIR principles make it a strategic enabler for interoperable digital ecosystems in both public and private sectors.

DCAT-AP (Application Profile) is a family of regional and domain-specific extensions of the core DCAT vocabulary, designed to meet the metadata requirements of specific communities—most notably within the European data ecosystem. These profiles build on



DCAT by introducing controlled vocabularies, mandatory fields, and validation rules tailored to local or sectoral needs.

With the release of DCAT v3, DCAT-AP profiles such as DCAT-AP v3, GeoDCAT-AP, and StatDCAT-AP have begun aligning with the new features and structural improvements of the core vocabulary. This includes:

- Support for dataset versioning and series, enabling better lifecycle management.
- Checksum and integrity metadata, improving trust and validation.
- Improved linking between datasets and services, enhancing discoverability.
- Relaxed cardinality and inverse properties, allowing more flexible modeling.

DCAT-AP profiles serve as semantic bridges between national or domain-specific metadata standards and the broader Linked Data ecosystem. By conforming to DCAT v3, these profiles ensure semantic interoperability across borders and sectors, facilitating cross-catalog search, data federation, and reuse in line with FAIR principles.

In practice, DCAT-AP profiles are widely adopted in open data portals, statistical offices, and geospatial infrastructures, and are supported by validation tools, SHACL shapes, and harmonized publishing guidelines.

Table 2 comparison table summarizing how DCAT-AP v3 extends and specializes DCAT v3 across key metadata dimensions

Dimension	DCAT v3 (Core Vocabulary)	DCAT-AP v3 (Application Profile)
Purpose	Generic metadata model for datasets and services	Tailored for European public sector interoperability
Scope	Global, cross-domain	European, with national and domain-specific extensions
Versioning Support	Introduces <code>dcat:version</code> , <code>dcat:previousVersion</code> , etc.	Mandates versioning metadata for datasets where applicable
Dataset Series	Defines <code>dcat:DatasetSeries</code>	Requires linking datasets to series for time-based or thematic grouping
Checksum & Integrity	Uses SPDX terms for checksum validation	Encourages checksum metadata for distributions
Controlled Vocabularies	Optional use of external vocabularies	Requires specific vocabularies (e.g., EuroVoc, ADMS)
Mandatory Fields	Minimal constraints	Defines mandatory and recommended fields for compliance
Interoperability Focus	RDF-based, Linked Data principles	Semantic harmonization across EU Member States



Validation Tools	SHACL shapes available	Provides SHACL validation rules and publishing guidelines
Profiles & Extensions	Extensible via RDF	Includes GeoDCAT-AP, StatDCAT-AP, DCAT-AP.de, etc.
Serialization Formats	RDF/XML, Turtle, JSON-LD	JSON-LD preferred for web publishing

The European Landscape of Semantic Interoperability

The EU-level agenda for promoting semantic (and other) interoperability is being advanced 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.

3.4 Open Digital Rights Language (ODRL)

ODRL is a policy expression language standardized by the W3C in 2018. It provides a flexible, interoperable information model and vocabulary for expressing permissions, prohibitions, and obligations related to digital content and services. Originally developed in 2000 to support Digital Rights Management (DRM) in mobile media, ODRL has evolved into a general-



purpose rights Scaling data spaces - Evidence, alignment, and emerging global convergence.docx

expression language used across sectors, including publishing, news, education, and financial services.

The language was first adopted by the Open Mobile Alliance (OMA) and later by the International Press and Telecommunications Council (IPTC) for RightsML. It became a W3C Community Group in 2011, ensuring transparent governance and long-term stability 1.

ODRL defines a core policy model consisting of:

- Permissions: Actions allowed on an asset (e.g., play, print).
- Prohibitions: Actions explicitly disallowed.
- Duties: Obligations that must be fulfilled for permissions to apply (e.g., payment).
- Constraints: Conditions like time, location, or usage limits.
- Parties: Entities involved (assigner, assignee).
- Assets: The digital resources being governed.

Policies are expressed in machine-readable formats (JSON-LD, XML, Turtle) and can be linked to content or embedded in metadata. The model supports policy inheritance, conflict resolution, and profile extensions for domain-specific needs.

ODRL has been widely adopted in mobile ecosystems, with over a billion devices supporting ODRL-based DRM 1. It is also used in news syndication, eBook licensing, Creative Commons, and market data licensing (e.g., the W3C Market Data Profile for ODRL) 3. The language's extensibility has enabled the creation of community-specific profiles, such as RightsML and ODRL for financial instruments, supporting automated compliance and rights enforcement in complex data supply chains.

ODRL is a semantic policy language grounded in Linked Data principles, enabling machine-readable, interoperable rights expressions. It plays a critical role in data sovereignty and policy-driven data sharing in data spaces and federated ecosystems. By providing a shared vocabulary and formal structure for expressing usage conditions, ODRL facilitates automated reasoning, compliance checking, and cross-platform policy enforcement.

However, the lack of formal semantics has been a challenge. Recent research is addressing this by developing declarative semantics for ODRL to ensure consistent interpretation across systems. This is essential for semantic interoperability, where different agents must interpret and act on policies in a uniform way.

3.5 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.

3.5.1 Common semantic data models

The Data Space Instance is responsible for ‘standardizing’ and developing common semantic data models within the data space instance. The Data Space Instance may use any means to introduce 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 with the rules and policies of this Data Space Instance. Finally, the Data Space Instance is responsible for supporting the domain-specific data models. Those may be managed with different governance models.

3.5.2 Semantic management data services

Semantic Management Data services operate within data space instances to facilitate the **alignment** of diverse semantic models with the agreed-upon common semantic data model. However, direct mapping may not always be the appropriate methodology — particularly in domains like cybersecurity, where such approaches could introduce ambiguity or compromise precision.

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. The semantic management infrastructure could be enhanced by Vocabulary Hubs, which support the publication, versioning, and provisioning of semantic models. A Data Space Instance has flexibility in its approach—it may operate one or multiple internal vocabulary hubs or leverage references to existing external instances, depending on the specific requirements and governance model of the data space.



4. Selection of best practices and good examples

Following the general motivation for semantic interoperability and the approach for Data Spaces in general, this section will illustrate such scenarios and insights on the current developments by making use of real-life use cases. The selected use cases span across different domains. Use cases need to follow the template presented below, and can be added via the IDSA Working Group Architecture.

Proposal for a template for use case description:

Name:

Description:

Type of interoperability needed by the
use case (--> EIF):

Reason:

Technical interoperability

Semantic interoperability

Organizational interoperability

Legal interoperability

Semantic interoperability scenarios
needed for realization of the use case
(--> Chapter 2):

Reason:

Understanding

Finding

Updating

Operating



5. Outlook

In summary, this paper has delved into the crucial concept of semantic interoperability within the field of Data Spaces, aligning it with the European Interoperability Framework and the facets outlined in ISO/IEC 21823-1:2019, respectively in ISO/IEC 19941:2017 .

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. The Dataspace Protocol focuses on the tech technical interoperability aspects, but builds on key approaches to semantic interoperability, DCAT v3, and ODRL.

In the concluding chapter, will provide an 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, 2023 and 2024.

These workshops, documented and detailed on the webpage semantic.internationaldataspaces.org¹⁴, 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 in 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

¹⁴ <https://semantic.internationaldataspaces.org/>



collaboration, these initiatives aim to guide and shape the future of interoperable data ecosystems, ensuring accessibility, transparency, and efficiency for all stakeholders involved.

The need for standardization in the field of semantic interoperability was recognized by the European Commission and the Standardisation Request [614 “European Trusted Data Framework”](#) contains two standardization deliverables to be found in the [annex](#).

Item 4: Data catalogue implementation framework

This (these) technical specification(s), entry 4 in table 1 of Annex I, shall provide a framework for standardised catalogue metadata, in support of the findability of data within and across data spaces.

The development of the technical specification(s) shall take into account the Interoperable Europe solutions based on the W3C Data Catalogue Vocabulary (DCAT) standard, in particular the DCAT-AP, DCAT-AP-HVD and GeoDCAT-AP profiles.

The technical specification(s) shall:

- a) set out the common catalogue metadata, to be applied across all common European data spaces;
- b) establish rules on the setting out of domain-specific catalogue metadata, to be applied in selected common European data spaces;

Item 5: Semantic assets implementation framework

This (these) technical specification(s), entry 5 in table 1 of Annex I, shall provide a framework for common, open vocabularies, classification schemes, taxonomies, code lists and ontologies, in support of the interpretation and analysis of shared data within and across data spaces.

Existing standards-based frameworks, in particular the Core Vocabularies and the Asset Description Metadata Schema Application Profile/ADMS-AP (EC - SEMIC), the Asset Administration Shell (IEC), the ISO/IEC 19115 metadata standards, the European Commission countries and territories reference data asset⁶ for geospatial data, and the Smart Applications REference/SAREF Ontology (ETSI), shall be taken into account.

The technical specification(s) shall:

- a) specify criteria for the selection of semantic assets;
- b) specify methods for the semantic annotation of shared data, the detailed metadata, based on the semantic assets mentioned under point (a).

6. 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://github.com/International-Data-Spaces-Association/IDS-RAM_4_0
- [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://iec.ch/basecamp/semantic-interoperability-challenges-digital-transformation-age>

CONTACT

International Data Spaces Association

Emil-Figge-Str. 80
44227 Dortmund | Germany

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

WWW.INTERNATIONALDATASPACE.ORG



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