Representing Knowledge in Dataspaces

Paul Moosmann[1], Rohit A. Deshmukh[1], Christoph Lange[1,2] and Johannes Theissen-Lipp[1,2]

[1]Fraunhofer Institute for Applied Information Technology FIT, Germany
[2]RWTH Aachen University, Germany

Abstract

Dataspaces are increasingly critical for enabling collaboration among diverse participants (individuals, institutions, or machines) in distributed environments. However, the effective sharing of data, knowledge, and services depends on a mutual understanding, which often remains poorly defined. Existing approaches focus on individual aspects of knowledge representation, but lack a holistic framework tailored to the requirements of dataspaces. This gap hinders seamless integration and mutual understanding across diverse participants, of both dataspace-specific and domain-specific content.

This paper addresses these challenges by providing a clear definition of knowledge in the scope of dataspaces. We identify answers to key questions regarding why, what, where, and how knowledge needs to be represented, and provide an overview of essential knowledge representations needed for acting in dataspaces. In addition, we examine existing solutions from the Semantic Web and building blocks from the Dataspaces Support Centre (DSSC) Blueprint, highlighting gaps and opportunities for improvement. Our contributions aim to provide an overview of knowledge representations in dataspaces and a systematic foundation for future research and development in this area.

Keywords

Dataspaces, Data Spaces, Knowledge Representation, Semantic Interoperability, Mutual Understanding

This paper addresses (Issue #8) of the W3C Dataspaces Community Group.

1. Introduction

Dataspaces [1, 2, 3] have emerged as a key paradigm for managing diverse and distributed data, offering flexible data sharing without requiring unification or centralization. The promise of *data sovereignty*, often referred to as *power to control [own] data* even when sharing it with others, is a major advantage of dataspaces over other data sharing approaches. Dataspaces thereby facilitate collaboration among diverse participants, which may include machines, software agents, or humans acting as end users or on behalf of institutions [2]. To operate effectively, all participants of a dataspace must share a mutual understanding of the data, knowledge, and services to be exchanged. This mutual understanding enables seamless interaction and supports interoperability by storing knowledge in the metadata of the actual data that is exchanged in a dataspace. Despite advances in knowledge representation in areas such as the Semantic Web or building blocks of the Dataspaces Support Centre (DSSC) Blueprint [4], the concept of "knowledge" in dataspaces remains poorly defined. Existing approaches often focus on individual components, such as specific use cases or isolated applications, schema matching, or ontology alignment. These approaches are not properly addressing the broader requirements for holistic knowledge representation across the entire dataspace ecosystem.

This lack of clarity creates several challenges. First, there is no consensus on what constitutes knowledge in dataspaces, making it difficult to standardize or evaluate methods for representing it. Second, there is limited guidance on where or how knowledge should be represented. Finally, current frameworks often overlook gaps in knowledge representation tools and building blocks that are critical for ensuring seamless collaboration and data sharing among dataspace participants. Addressing these

The Third International Workshop on Semantics in Dataspaces, co-located with the Extended Semantic Web Conference, June 01, 2025, Portorož, Slovenia
paul.moosmann@fit.fraunhofer.de (P. Moosmann); rohit.deshmukh@fit.fraunhofer.de (R. A. Deshmukh); christoph.lange-bever@fit.fraunhofer.de (C. Lange); theissen-lipp@dbis.rwth-aachen.de (J. Theissen-Lipp)
0009-0005-2114-8578 (P. Moosmann); 0000-0003-2885-7076 (R. A. Deshmukh); 0000-0001-9879-3827 (C. Lange); 0000-0002-2639-1949 (J. Theissen-Lipp)

gaps is essential for realizing the full potential of dataspaces in fulfilling their promises, such as data sovereignty, in complex or distributed systems.

This paper addresses these gaps by providing a clear definition of knowledge in the scope of dataspaces and identifying key considerations for its representation. We explore why, what, where, and how knowledge needs to be represented, and present an overview of essential representations required for dataspace operations. We also analyze existing Semantic Web artifacts and DSSC components, such as the DSSC Building Blocks, Dataspace Services, and Toolbox, identifying strengths, limitations, and opportunities for advancement. Through this work, we aim to provide a foundation for improving knowledge representation in dataspaces and to guide future research and practical implementations.

It is important to note that this paper focuses on declarative knowledge, which captures structured facts, rules, constraints, and relationships within dataspaces. In contrast, procedural knowledge – concerned with workflows, processes, and sequences of actions – is not included, as it is highly context-dependent and often requires specialized execution mechanisms beyond static knowledge representation. Therefore, procedural knowledge is beyond the scope of this paper.

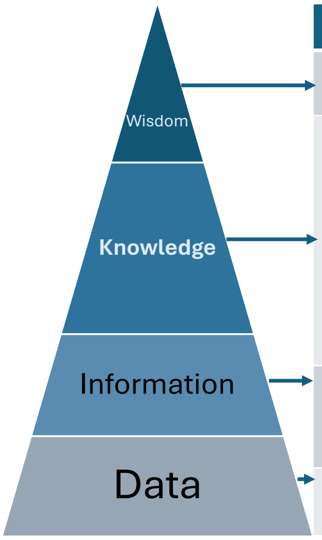
2. Knowledge in Dataspaces

Current dataspace initiatives are developing design principles, architectures, and technologies for dataspaces. For instance, the DSSC is developing a *Blueprint* that defines terminologies and identifies building blocks for dataspaces [4]. At this stage in the evolution of dataspaces, defining the concept of knowledge is an essential prerequisite to developing standards and technologies that facilitate its effective management. In addition, some initiatives are working towards making dataspaces FAIR (Findable, Accessible, Interoperable, and Reusable) [5, 6], and having a clear definition of knowledge can help in streamlining and harmonizing their activities.

The ultimate goal of creating and using knowledge is to enable effective decision-making with minimal risk [7] which also applies to dataspaces. For example, in an industrial setting, predictive maintenance for factory machines relies on knowledge derived from historical patterns and real-time data, such as temperature, humidity, vibrations, and pressure. By leveraging this knowledge, potential failures can be anticipated, allowing for timely interventions and minimizing operational disruptions.

In computer science, hierarchical models such as the Data-Information-Knowledge-Wisdom (DIKW) pyramid [8] are commonly used to illustrate the relationships between raw data, structured information, contextualized knowledge, and ultimately, wisdom. We adapt this DIKW model to dataspaces to explore how data, through the addition of various forms of semiotics (syntax, semantics, and pragmatics) [9], progresses to higher levels of abstraction, ultimately transforming raw data into knowledge that supports decision-making. The table in Figure 1 maps different aspects of semiotics – syntax, semantics, and pragmatics – to the layers of the DIKW pyramid in the context of dataspaces. Data (D) consists of raw values, their data types, and their representation using basic syntax. Information (I) emerges when structured, machine-readable syntax (e.g., JSON or XML) and contextual semantics are introduced, enabling interpretability but allowing for multiple possible meanings. Knowledge (K) is formed by enriching information with formal specifications of concepts, their semantics and relationships, application-specific constraints, business rules, reasoning mechanisms, historical patterns, and pragmatics, making it actionable for decision-making. Wisdom (W) extends knowledge by evaluating and refining decisions, assessing their correctness, efficiency, and long-term impact, ultimately enabling continuous learning and optimization. Building on this perspective, we derive a working definition of knowledge in dataspaces as follows:

> *Knowledge in dataspaces is the state that emerges when data, progressively enriched with semiotics (syntax, semantics, and pragmatics), attains an unambiguous, machine-understandable form that contributes to informed decision-making while minimizing the risk of errors, inefficiencies, or unintended consequences in a given use case scenario. Knowledge can be represented as a complete framework or as modular components that support specific aspects.*



Layer	Semiotics (Syntax, Semantics, Pragmatics)	Example / Explanation *(Please read from bottom to top)*
W	• Decision Evaluation • Impact Analysis • Decision Refinement	**Evaluation and refinement** of the effectiveness of decisions such as 'screening only family-friendly movies on weekend afternoons' (e.g., assessing and improving correctness, efficiency, and long-term impact).
K	• Historical Data Patterns • Causal Semantics • Pragmatics • Business Rule and Policy Semantics • Reasoning/Inference Semantics • Application-specific Semantics • Ontological Semantics with Formal, Unambiguous, Machine-understandable Syntax	**Historical insights**: Historical data suggesting that 'screening of family-friendly movies is usually profitable on weekend afternoons' can be used to make decisions. **Cause and effect**: Higher ticket prices generally lead to lower attendance **Pragmatics:** Adjusting interpretations of rules based on real-world factors. E.g., Interpretation of profitability in relation to ticket rates and occupancy; interpretation of family-friendly movies based on country. **Business rule**: Screening at Theater A is profitable only if occupancy > 50. **Reasoning**: Subsumption Reasoning can used to infer that accessibility features modelled for a place can be applied to theatres because theatres are modelled as places. **Application profiles (constraints)**: Theater A has only one screen, with a max occupancy of 100. **Universal knowledge representation**: Formal specification of concepts (Theater, Address, Movie, Screening, Screen, Data/Time, Occupancy, Audience) and their relationships.
I	• Contextual Semantics • Property/Attribute Semantics • Machine-readable Object/Record Syntax	**Context**: 21 is the occupancy of Theater A at address B for movie C's screening on date/time D **Syntax**: Data format such as JSON, XML, or CSV, etc. *(Here, even if the syntax and context are known, the exact definitions of these concepts and their relationships are not explicitly defined. Multiple interpretations are possible.)*
D	• Data Type Semantics • Data Type Syntax • Raw Data	**Data type semantics**: Type interpretation: The number "21" denotes the quantity twenty-one. **Data type syntax:** How numbers are represented using symbols (digits, punctuation, +/- sign). **Raw data:** "21"

Figure 1: The DIKW pyramid [8] adapted to dataspaces

To materialize various layers of semantics illustrated in the table, dataspaces can leverage formal models and technologies developed by the Semantic Web community. The Semantic Web technology stack [10] offers standards and tools for structuring and reusing knowledge. Further research is required to determine which constructs – such as ontologies, application profiles, constraint validation shapes, or a combination thereof – are best suited for knowledge representation in dataspaces. For instance, the Web Ontology Language (OWL) [11] is based on Description Logic, a decidable fragment of First-Order Predicate Logic, making it a possible candidate for representing knowledge in dataspaces. While OWL ontologies and vocabularies are widely used for universal knowledge representation, applying them to dataspaces raises challenges, particularly in reconciling their open-world assumption with the need for an application-oriented representation [12]. The role of application profiles [13, 14] in addressing these challenges remains an active area of research. We further discuss the current state of technologies for knowledge representation in dataspaces in Section 4.

3. Requirements for Knowledge Representation in Dataspaces

To derive the requirements for knowledge representation in dataspaces, we consider the functional requirements that dataspaces must fulfill. In Section 4, we then map existing technologies to these requirements, highlighting how current solutions can support the identified knowledge representation needs. We survey existing dataspace specifications to establish a basis for our analysis. These specifications include the Gaia-X Architecture Model [15] and the IDS Reference Architecture Model (IDS RAM) [16]. While the Gaia-X Architecture Model focuses on a more constrained model, the IDS RAM provides a comprehensive overview of general functional requirements as well as specific instances of these requirements. We use the IDS RAM as the basis for our analysis because it is subdivided more finely when considering the technical requirements. The IDS RAM is a comprehensive model, and we focus on the first four (of six) functional layers [16]:

1. Trust
2. Security & Data Sovereignty
3. Ecosystem of Data
4. Standardized Interoperability

We decided to constrain the scope of our analysis to these four layers, since from our point of view, they represent the most fundamental technical aspects required to establish a dataspace. On

Table 1
Requirements for Knowledge Representation in Dataspaces

Existing functional requirements for dataspaces		Requirements for Knowledge Representation in Dataspaces	
Objective	**Core Functionality**	**What knowledge needs to be represented?**	**Examples**
Trust	Roles	(1) Dataspace User Roles such as Dataspace Participant (an Organization/ a Person on behalf of an Organization) who can take further roles such as Data Product (aka Offering / Resource / Asset) Owner, Data Product Provider,and Data Product Consumer, etc. [4] (2) Dataspace Administrator Roles such as Dataspace Governance Authority [4] (3) Dataspace Intermediary Roles are of three types: Core Intermediaries and an Operator, Participant Agent Intermediaries, and Personal Data Intermediaries [4]	(1) A company wanting to share year-wise temperature time series datasets for Berlin for the years 2010-2020. (2) A group of dataspace participants recognized by the dataspace that is responsible for developing, maintaining, and enforcing the internal governance rules. (3) A company providing Connector-as-a-Service.
	Identity Management	Specification of an identity scheme and a format, and assignment of identities to: (1) Dataspace Participants (Persons, Organizations, Groups of Persons or Organizations) (2) Data Product (aka Offering / Resource / Asset) (3) Connector (aka Participant Agent) (4) Various entities such as instances of semantic metadata models in knowledge graphs	There is a persistent identifier infrastructure for a dataspace that issues URIs based on a defined scheme for each participant, data product, and connector, etc.
	User Certification	Certification-related metadata: Conformity schema, Conformity criteria, Trust levels, Conformity/Compliance status, Attestation	A participant obtains an IDS certification for their Connector, verifying compliance with security standards.
Security and Data Sovereignty	Authentication and Authorization	Authentication: Same as "Identity Management" + proof of identity Authorization: Access & Usage Policies and their relationships with Dataspace participants, data products, etc.	(1) Connector A, wanting to access a data product from Connector B, presents an access token issued by the dataspace's Identity and Access Management (IAM) system. Connector B validates the token before granting access according to the defined access policies. (2) Access policy with a spatial restriction: Only consumers based in Germany can access this data product
	Usage Policies (and Usage Enforcement)	Data product usage policies. Contracts between the participants. Persistent links to participant identities.	(1) Temporal restrictions: Consumers can use this data product only until 31 Dec 2025. (2) Spatial restrictions: Consumers can use this data product only when their current location is Germany. (3) Restrictions on sharing: Consumers can use the data product but not share it with others.
	Trustworthy Communication and Security by Design	Different security profiles for the connectors.	A Connector enforces encrypted communication (e.g., TLS 1.3) for secure data exchanges.
	Technical Certification	Requirements for successful certification.	The governance body requires a certification for the connector from a certification body in order to establish trust among all participants.
Ecosystem of Data	Data Source Description	Semantic metadata model capable of describing the data source.	A dataspace governance authority mandates the use of DCAT for describing the metadata of datasets consistently across different providers.
	Brokering	Semantic models to describe the catalogue(s) of dataspace participants, data offerings, and service offerings and their interfaces.	A Metadata Broker maintains a catalog of available data products in the dataspace and offers a SPARQL-based query interface.
	Vocabularies	Knowledge modelled by the vocabularies. Metadata information of the vocabularies.	A dataspace governance authority recommends the use of Semantic Sensor Network (SSN) ontology to its participants for describing sensors and their observations.
Standardized Interoperability	Operation	Idendity of the user of the connector. Knowledge of data flow for accurate logging.	There has been an issue with the data exchange, so the logging has to be accessed. In case of a contract breach, the user of the connector needs to be identifiable.
	Data Exchange	See Data Source Description and Usage Policies.	(1) A data analysis service provider wants to access the data to perform their analysis, while the data provider wants to make sure that their data is only used according to the agreed upon policies. (2) A data consumer uses the API specification outlined in the Data Source Description – including protocol, access method (pull/push), payload, and update frequency – to invoke the data service provided by the data provider, and retrieves data.

the other hand, the dismissed layers focus on factors that add functionalities to a dataspace already in operation. For each core functionality within these four layers, we identified the knowledge that needs to be represented to enable the respective functionality. In addition, we provided example situations that might occur in real-world scenarios to illustrate how knowledge representation is necessary to implement the functionalities. The results of this analysis are summarized in Table 1, which systematically maps each core functionality to the corresponding knowledge requirements and example scenarios.

4. Existing Solutions supporting Knowledge Representation in Dataspaces

In this section, we assess how existing Semantic Web artifacts and dataspace components and services address the requirements for knowledge representation in dataspaces we derived in the previous section. We introduce key solutions relevant to dataspaces, each followed by an evaluation of their effectiveness in meeting these requirements.

4.1. Semantic Web Artifacts

We group artifacts from the areas of Semantic Web technologies and evolving Semantic Web standards into categories of application in dataspaces and discuss their effectiveness to address the requirements derived in the previous section:
Defining What Exists: Taxonomies, vocabularies, and ontologies (e.g., DCAT or Schema.org) provide structured ways to define entities and concepts within dataspaces. Taxonomies establish hierarchical relationships, vocabularies define controlled terms, and ontologies offer rich, formalized descriptions with logical constraints. These artifacts contribute to knowledge representation across all the functional

requirements listed in Table 1, since they provide or at least contribute to the information models needed to implement the functional requirements of dataspaces.

Specifying How to Use Things: Application profiles, particularly when expressed in SHACL (Shapes Constraint Language [17]) and ShEx (Shape Expressions [18]), enable the definition of constraints on the structure of data instances. These mechanisms ensure that data adheres to expected structures, supporting interoperability and validation across diverse dataspace participants. These artifacts are paramount for dataspace actions, where it is necessary to adhere to certain constraints. This is the case when creating semantic descriptions of dataspace participants, services, and resources, or when designing usage policies [19].

Linking Data with Other Data: Linked Data principles [20] provide mechanisms for interconnecting datasets across distributed environments. By utilizing RDF and URIs, Linked Data fosters semantic interoperability, enabling the integration of heterogeneous data sources. The Linked Data approach can be used for the data exchange by providing an identifier of a data offering that is dereferenceable as a URL where the data can be accessed. However, other solutions exist, such as listing the data offering in a metadata catalogue that contains the endpoint information of the connector through which it is accessible.

Representing and Integrating Knowledge: Knowledge Graphs [21, 22, 20] serve as a foundational technology for organizing, structuring, and linking information in dataspaces. They encapsulate semantic relationships between entities, supporting context-aware data access and reasoning [23]. In the context of dataspaces, knowledge graphs can be used to integrate data from heterogeneous data sources or to store the meta-data information of participants, services, and data offerings. It can be relevant for brokering, data source description, and data exchange.

Querying Knowledge: SPARQL [24] and GraphQL [25, 26] for RDF enable querying and retrieving knowledge from structured data sources. While SPARQL provides expressive querying capabilities based on graph patterns, GraphQL for RDF allows flexible and efficient data retrieval tailored to user needs. These artifacts are relevant for dataspaces in order to ensure that data and service offerings can be queried in order for participants to effectively find what they are looking for. In the functional dataspace requirements, this is particularly relevant to the brokering aspect.

Identity, Trust, and Provenance: The PROV-O (Provenance Ontology [27]) model supports capturing, representing, and reasoning over provenance information. This ensures transparency, accountability, and trustworthiness of data exchanges in dataspaces. The Verifiable Credentials Data Model [28] provides a solution for tamper-evident credentials whose authorship can be cryptographically verified. Decentralized identifiers (DIDs) are a type of identifier that enables a verifiable, decentralized digital identity. DIDs have been designed so that they may be decoupled from centralized registries, identity providers, and certificate authorities [29]. These artifacts provide solutions to fulfill the functional requirements of dataspaces regarding trust, security, and data sovereignty.

Policy Representation and Enforcement: The Open Digital Rights Language (ODRL) [30] provides a policy expression framework for defining access control and usage permissions. Its role in dataspaces is critical for enforcing governance and regulatory compliance. This artifact helps with the implementation of usage policies attached to a data exchange action. This helps to achieve data sovereignty, a key requirement for dataspaces.

Mapping Heterogeneous Data to RDF: RML (RDF Mapping Language) [31, 32] facilitates the transformation of diverse data formats into RDF. This mapping process is essential for harmonizing structured and semi-structured data sources within dataspaces [33]. This artifact is relevant to the goal of dataspaces of exchanging data from heterogeneous data sources. After mapping the data to RDF it can be stored in a knowledge graph, as described in *Representing and Integrating Knowledge* above.

The existing Semantic Web artifacts could fill an important role in enabling knowledge representation in dataspaces. They enable structured data representation, interoperability, and governance through taxonomies, ontologies, and Linked Data principles. Knowledge Graphs and RDF mappings facilitate data integration from heterogeneous sources, while query languages enable discoverability. While several artifacts exist to help enable knowledge representation, they often serve only as a starting point and do not necessarily provide all required functionalities. E.g., existing ontologies or vocabularies can be used to create models to represent knowledge, but their capabilities might be limited, especially when aiming to model very specific domain knowledge. Extending Semantic Web artifacts (and even using existing ones) poses a challenge in most cases, since not everyone is proficient in the use of Semantic Web technologies. So, while there are possibilities to assist in the implementation of the functional dataspace requirements, using them is often only done reluctantly since their high complexity poses an entry threshold. Therefore, further research is needed regarding user-friendly tooling assisting with the use of Semantic Web technologies, to enable the broader adoption of Semantic Web technologies.

4.2. Dataspace Components and Services

This section identifies functional building blocks and components that leverage Semantic Web artifacts to represent knowledge in dataspaces.

The DSSC project's partners, along with its extensive network of stakeholders, include all major dataspace development and deployment initiatives, as well as key dataspace practitioners in Europe. As a result, the DSSC Blueprint [4] serves as a convergence point for dataspaces while also advancing their maturity [34]. We therefore use it as a de facto reference point for assessing the current dataspace components and tools in relation to knowledge representation.

The DSSC Blueprint [4] defines business, organizational, and technical building blocks for dataspaces. Business and organizational building blocks define knowledge elements such as business models, governance rules (rulebook), roles and responsibilities of administrators, users, and intermediaries, as well as legal frameworks for dataspaces. Technical building blocks define functional requirements related to data interoperability, sovereignty, trust, and data value creation enablers. The Blueprint recommends the use of foundational standards such as W3C Verifiable Credentials, DCAT v3, and ODRL for representing associated knowledge. The requirements outlined by the building blocks are realized through Dataspace Services [35], which fall into three high-level categories:

Federation Services: Provide infrastructure components that facilitate the discovery and interaction of participants and their data products. These services represent knowledge such as dataspace user roles, identities, certifications, access and usage policies, and semantic metadata models (ontologies, vocabularies, and application profiles).

Participant Agent Services: Enable participants to interface with a dataspace, share their data, and attach policies to their data products. These services instantiate knowledge models, including user roles, identities, certifications, metadata models, and policies.

Value Creation Services: Facilitate value generation on top of shared data, instantiate business models, and enable data valorization. These services combine various types of knowledge to support decision-making. Examples include data marketplaces, data analytics services, and training and education services.

These categories encompass various conceptual components, enabling flexible implementations. The DSSC Toolbox [36] offers a public catalogue of dataspace tools, curated by the DSSC to support these functionalities. While these technical frameworks provide a solid foundation, significant challenges remain in the implementation and adoption of knowledge representation artifacts and approaches within dataspaces. Even six months after its release, the DSSC Toolbox lists only 20 tools as of March 2025. Other dataspace tools remain at various Technology Readiness Levels and are often difficult to discover. Some widely used tools, such as the Eclipse Dataspace Connector (EDC) [37], lack support for proper representation of semantic models [38], while the XFSC Federated Catalogue [39], a Metadata Broker implementation, lacks an RDF-based backend, limiting its ability to support SPARQL queries [40].

Beyond the availability of tools, dataspaces also face significant usability challenges that directly impact knowledge representation. The Semantic Web is often perceived as complex by developers and domain experts [12], requiring expert-level capabilities for onboarding and interfacing with dataspaces. There is a pressing need for user-friendly, open-source tools such as collaborative Semantic Web IDEs (Integrated Development Environments), as well as Generative AI-powered assistants that can aid in the creation of semantic models by leveraging existing ontologies, vocabularies, and application profiles [12].

Additionally, application profiles – which play a crucial role in representing semantic models, enforcing closed-world assumptions in dataspaces, and enabling real-world use cases – remain underexplored. Open research questions persist regarding formats, technologies, and practical implementation strategies for application profiles. Thus, further research is needed to advance tooling, application-specific knowledge representation technologies, and user-friendly interfaces for knowledge representation and exploitation in dataspaces. Addressing these gaps is essential for enabling effective decision-making with minimal risk.

5. Summary and Outlook

In this paper, we highlighted the importance of knowledge for enabling informed decision-making in dataspaces and systematically identified requirements for knowledge representation. First, we outlined key gaps, including the lack of a clear definition of knowledge in dataspaces and the fact that existing tools, technologies, building blocks, and frameworks each follow their own implicit understanding of knowledge. This lack of standardization leads to interoperability challenges, inconsistencies, and incomplete knowledge representation. Second, we illustrated how knowledge emerges by enriching raw data and structured information with various forms of semiotics (syntax, semantics, and pragmatics). We mapped different types of semiotics to the layers of the DIKW pyramid and presented a working definition of knowledge in the context of dataspaces. Third, we analyzed the functional requirements for dataspaces as defined by the IDS RAM and systematically identified the relevant types of knowledge needed to fulfill each requirement. Finally, we examined how existing Semantic Web artifacts, dataspace components, services, and tools can support knowledge representation and identified critical gaps that hinder effective integration and knowledge sharing within dataspaces. In future work, we plan to investigate the role of application profiles in representing semantic models, enforcing closed-world assumptions, and enabling real-world use cases in dataspaces. More importantly, future work will involve investigating whether the DSSC blueprint outlines additional functional requirements when compared to the IDS-RAM we referenced, followed by a more fine-grained mapping of the existing knowledge representation solutions discussed in sections 4.1 (Semantic Web artifacts) and 4.2 (DSSC components) to the DIKW pyramid introduced in section 2, as well as a systematic evaluation of these solutions based on this mapping. At a higher level, a long-term goal is to develop easy-to-use tools for intuitive semantic model building and to explore generative AI as a means to increase productivity, simplify user interaction, and improve engagement in knowledge creation and representation tasks.

Acknowledgments

Funded by the Deutsche Forschungsgemeinschaft under Germany's Excellence Strategy – EXC-2023 Internet of Production – 390621612, the Data Spaces Support Centre project within the European Union Digital Europe Programme, under grant agreement number 101083412, and supported by a Fraunhofer ICON grant through the Next Generation Dataspaces Initiative.

Declaration on Generative AI

During the writing of this paper, the author(s) used DeepL and GPT-4o in order to: Grammar, translation and spelling check. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] A. Halevy, M. Franklin, D. Maier, Principles of Dataspace Systems, in: Proceedings of the Twenty-Fifth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS '06,

Association for Computing Machinery, New York, NY, USA, 2006, p. 1–9. doi:10.1145/1142351.1142352.
[2] B. Otto, M. ten Hompel, S. Wrobel, International data spaces, in: Digital Transformation, Springer Berlin Heidelberg, Berlin, Heidelberg, 2019, pp. 109–128. doi:10.1007/978-3-662-58134-6_8.
[3] J. Theissen-Lipp, M. Kocher, C. Lange, S. Decker, A. Paulus, A. Pomp, E. Curry, Semantics in Dataspaces: Origin and Future Directions, in: Companion Proceedings of the ACM Web Conference 2023, WWW '23 Companion, Association for Computing Machinery, New York, NY, USA, 2023, p. 1504–1507. doi:10.1145/3543873.3587689.
[4] Data Spaces Support Centre, Data Spaces Blueprint v1.5, Technical Report, 2024. URL: https://dssc.eu/space/bv15e/766061169/Data+Spaces+Blueprint+v1.5+-+Home, accessed on 2025-02-25.
[5] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, et al., The FAIR guiding principles for scientific data management and stewardship, Scientific Data 3 (2016). doi:10.1038/sdata.2016.18.
[6] M. Hauff, L. M. Comet, P. Moosmann, C. Lange, I. Chrysakis, J. Theissen-Lipp, FAIRness in Dataspaces: The Role of Semantics for Data Management, in: The Second International Workshop on Semantics in Dataspaces, co-located with the Extended Semantic Web Conference, 2024.
[7] A. Aamodt, M. Nygård, Different roles and mutual dependencies of data, information, and knowledge—an ai perspective on their integration, Data & Knowledge Engineering 16 (1995) 191–222.
[8] J. Rowley, The wisdom hierarchy: representations of the dikw hierarchy, Journal of information science 33 (2007) 163–180.
[9] S. Decker, Semantic web methods for knowledge management, Ph.D. thesis, Karlsruhe, Univ., Diss., 2002, 2002.
[10] O. Signore, et al., Representing knowledge in the semantic web, in: Open Culture Conference (organised by the Italian office of W3C), 2005, pp. 27–29.
[11] J. Bao, D. Calvanese, B. C. Grau, M. Dzbor, A. Fokoue, C. Golbreich, S. Hawke, I. Herman, R. Hoekstra, I. Horrocks, E. Kendall, M. Krötzsch, C. Lutz, D. L. McGuinness, B. Motik, J. Pan, B. Parsia, P. F. Patel-Schneider, S. Rudolph, A. Ruttenberg, U. Sattler, M. Schneider, M. Smith, E. Wallace, Z. Wu, A. Zimmermann, J. Carroll, J. Hendler, V. Kashyap, OWL 2 Web Ontology Language, W3C Recommendation, W3C OWL Working Group, 2012. URL: https://www.w3.org/TR/owl2-overview/, accessed on 2024-05-23.
[12] R. A. Deshmukh, D. Collarana, J. Gelhaar, J. Theissen-Lipp, C. Lange, B. T. Arnold, E. Curry, S. Decker, Challenges and Opportunities for Enabling the Next Generation of Cross-Domain Dataspaces, in: The Second International Workshop on Semantics in Dataspaces, co-located with the Extended Semantic Web Conference, 2024.
[13] R. Heery, M. Patel, Application profiles: mixing and matching metadata schemas, Ariadne 25 (2000).
[14] SEMIC, Application Profiles: What are they and how to model and reuse them properly? A look through the DCAT-AP example., 2023. URL: https://europa.eu/!DmBHvH.
[15] Gaia-X European Association for Data and Cloud, Gaia-X Architecture Document - 24.04 Release, Technical Report, 2024. URL: https://docs.gaia-x.eu/technical-committee/architecture-document/24.04/, accessed on 2025-03-06.
[16] B. Otto, S. Steinbuß, A. Teuscher, S. Bader, Reference Architecture Model Version 4.0, Technical Report, International Data Spaces Association, 2022. URL: https://docs.internationaldataspaces.org/ids-ram-4/.
[17] H. Knublauch, D. Kontokostas, Shapes constraint language (SHACL), W3C Recommendation, W3C Working Group, 2017. URL: https://www.w3.org/TR/shacl/, accessed on 2024-05-23.
[18] E. Prud'hommeaux, J. E. Labra Gayo, H. Solbrig, Shape expressions: an RDF validation and transformation language, in: SEMANTICS, 2014, pp. 32–40. doi:10.1145/2660517.2660523.
[19] P. Moosmann, J. Theissen-Lipp, C. Lange, Enhanced and Scalable RDF Validation Techniques for Dataspaces, in: International Conference on Semantic Systems, 2024.
[20] C. Bizer, T. Heath, T. Berners-Lee, Linked Data - The Story So Far, Int. J. Semantic Web Inf. Syst. 5

(2009) 1–22.
[21] E. W. Schneider, Course Modularization Applied: The Interface System and Its Implications For Sequence Control and Data Analysis., Technical Report, 1973.
[22] A. Singhal, et al., Introducing the knowledge graph: Things, not strings, Official Google Blog 5 (2012) 3. URL: https://blog.google/products/search/introducing-knowledge-graph-things-not/.
[23] H. Paulheim, Knowledge graph refinement: A survey of approaches and evaluation methods, Semantic web 8 (2017) 489–508.
[24] C. B. Aranda, O. Corby, S. Das, L. Feigenbaum, P. Gearon, B. Glimm, S. Harris, S. Hawke, I. Herman, N. Humfrey, N. Michaelis, C. Ogbuji, M. Perry, A. Passant, A. Polleres, E. Prud'hommeaux, A. Seaborne, G. T. Williams, SPARQL 1.1 Query Language for RDF, W3C Recommendation, The W3C SPARQL Working Group, 2013. URL: https://www.w3.org/TR/sparql11-overview/, accessed on 2025-02-25.
[25] Facebook, Inc. (2015-2018), GraphQL contributors (2019-present), GraphQL (October 2021 Edition), Specification, The GraphQL Foundation, 2021. URL: https://spec.graphql.org/October2021/, accessed on 2025-02-25.
[26] R. Taelman, M. Vander Sande, R. Verborgh, Bridges between graphql and rdf, in: W3C Workshop on Web Standardization for Graph Data. W3C, 2019, pp. 4–7.
[27] T. Lebo, S. Sahoo, D. McGuinness, K. Belhajjame, J. Cheney, D. Corsar, D. Garijo, S. Soiland-Reyes, S. Zednik, J. Zhaos, PROV-O: The PROV Ontology, W3C Recommendation, The W3C SPARQL Working Group, 2013. URL: https://www.w3.org/TR/prov-o/, accessed on 2024-05-23.
[28] M. Sporny, D.Longley, D. Chadwick, I. Herman, Verifiable Credentials Data Model, 2025. URL: https://www.w3.org/TR/vc-data-model-2.0/.
[29] M. Sporny, D. Longley, M. Sabadello, D. Reed, O. Steele, C. Allen, Decentralized Identifiers (DIDs) v1.0, 2022. URL: https://www.w3.org/TR/did-1.0/.
[30] R. Iannella, M. Steidl, S. Myles, V. Rodriguez-Doncel, Odrl vocabulary & expression 2.2: W3c recommendation, 15 february 2018, 2018. URL: https://www.w3.org/TR/odrl-vocab/.
[31] A. Dimou, M. Vander Sande, P. Colpaert, R. Verborgh, E. Mannens, R. Van de Walle, RML: A Generic Language for Integrated RDF Mappings of Heterogeneous Data, in: Ldow, 2014.
[32] A. Dimou, M. Vander Sande, P. Colpaert, R. Verborgh, E. Mannens, R. Van de Walle, RDF Mapping Language (RML), W3C, Unofficial Draft 15 (2020).
[33] J. Theissen-Lipp, N. Schäfer, M. Kocher, P. Hochmann, M. Riesener, S. Decker, Extending Semantic RML Mappings with Additional Source Formats, in: Proceedings of the 26th International Conference on Enterprise Information Systems - Volume 1: ICEIS, INSTICC, SciTePress, 2024, pp. 201–208. doi:10.5220/0012551300003690.
[34] M. Bacco, A. Kocian, S. Chessa, A. Crivello, P. Barsocchi, What are data spaces? systematic survey and future outlook, Data in Brief 57 (2024) 110969.
[35] Data Spaces Support Centre, Data Spaces Blueprint v1.5 - Services for Implementing Technical Building Blocks, Technical Report, 2024. URL: https://dssc.eu/space/bv15e/766067344/Services+for+Implementing+Technical+Building+Blocks, accessed on 2025-03-06.
[36] Data Spaces Support Centre, DSSC Toolbox, 2025. URL: https://toolbox.dssc.eu/.
[37] Eclipse Foundation and community, Eclipse Dataspace Connector (EDC), 2024. URL: https://github.com/eclipse-edc/Connector.
[38] V. Alexiev, Keynote: Semantic problems in dataspaces, Presented at the AIOTI Workshop on Semantic Interoperability for Digital Twins, 2025. URL: https://bscw.ercim.eu/pub/bscw.cgi/d1321125/Vladimir-Alexiev.pdf, slides available at https://bscw.ercim.eu/pub/bscw.cgi/d1321347/Vladimir-Alexiev.pdf.
[39] Eclipse Foundation and community, XFSC Catalogue, 2023. URL: https://gitlab.eclipse.org/eclipse/xfsc/cat/fc-service.
[40] B. T. Arnold, K. Baydoun, D. Collarana, S. Duda, C. Gillmann, A. Hemid, P. Hertweck, P. Moosmann, D. Sukhoroslov, C. Lange, XFSC: A Catalogue of Trustable Semantic Metadata for Data Services and Providers, arXiv preprint arXiv:2501.14473 (2025).