

Adapting Ontology-based Data Access for Data Spaces

Medina Andresel¹, Veronika Siska¹, Robert David², Sven Schlarb¹ and Axel Weißenfeld¹

¹AIT Austrian Institute of Technology GmbH, Vienna, Austria

²Semantic Web Company GmbH, Vienna, Austria

Abstract

The exponential growth of data across various sectors requires robust frameworks for efficient data management and exchange, particularly in the context of training and improving artificial intelligence (AI) models. Data spaces emerge as a solution, facilitating seamless data exchange among organizations while safeguarding data sovereignty. This article explores the landscape of data spaces, emphasizing the role of Semantic Web standards in achieving interoperability and facilitating data sharing. It begins with use cases in crisis management and manufacturing to provide concrete requirements for discussing data space challenges and benefits. The IDS Data Space Architecture is presented, alongside an examination of the relevance of Semantic Web standards for data sharing. Examples of searching using Ontology-based Data Access (ODBA) offer insights into the potential of Semantic Web technologies to further improve the interoperability within data spaces. Finally, we explore how to setup a data space and publish data to enable OBDA-based search and the process to conduct the search itself.

Keywords

Semantic Web Technologies, Data Spaces, Ontology-based Data Access, Metadata

1. Introduction

Data is the driving factor for numerous IT systems in practical applications, ranging from curated enterprise data utilized for informed business decisions to training data necessary for refining machine learning algorithms and enhancing artificial intelligence (AI) models. Since various use cases and systems rely on extensive data repositories, the necessity arises to facilitate data sharing, oftentimes driven by commercial interests, all while preserving data sovereignty. Hence, there is a need for robust data management and exchange frameworks.

This is where the concept of data spaces [1] offers a framework for enabling efficient data exchange between organisations while preserving the data sovereignty for each participating entity. This article outlines the landscape of data spaces with a focus on how Semantic Web technologies can be used to achieve interoperability and enable data sharing between partner organisations and systems. In particular, we propose to enhance data spaces with the functionalities that the *Ontology-based Data Access* (OBDA) paradigm [2] offers. In OBDA, access to


The Second International Workshop on Semantics in Dataspaces, co-located with the Extended Semantic Web Conference, May 26 – 27, 2024, Hersonissos, Greece

✉ Medina.Andresel@ait.ac.at (M. Andresel); Veronika.Siska@ait.ac.at (V. Siska); robert.david@semantic-web.com (R. David); Sven.Schlarb@ait.ac.at (S. Schlarb); Axel.Weissenfeld@ait.ac.at (A. Weißenfeld)

🆔 0009-0002-4424-7817 (M. Andresel); 0000-0002-8057-1203 (V. Siska); 0000-0002-3244-5341 (R. David); 0000-0003-3717-0014 (S. Schlarb); 0000-0002-7246-2744 (A. Weißenfeld)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

disparate and heterogeneous data sources is mediated by an ontology through the mappings of raw information to concepts and relations defined in the domain-specific ontology. In OBDA, access to all data sources is then enabled by means of standard SPARQL [3] querying.

Our main contribution is a conceptual model that integrates and adapts OBDA to the requirements imposed by the data protection and sharing mechanisms existing in data spaces. The outline of this paper is as follows. We start by presenting use cases in two key areas: supporting authorities in crisis situations and manufacturing. This will serve as a basis for discussing the challenges and benefits associated with data spaces. We then proceed by introducing the IDS Data Space Architecture and outline the role of Semantic Web standards within this framework. We continue with our conceptualization of how OBDA can be realized in the data space in relation to the use cases described before. We then delve into data space setup and explain the principle of publishing and accessing data within this novel framework. Finally, we highlight discussion points and challenges and provide our conclusions.

2. Preliminaries

In this section we present two motivating use-cases, and briefly describe the concept of data spaces and existing Semantic Web based approaches used for data spaces.

2.1. Use Cases

2.1.1. Supporting Authorities in Crisis Situations

In the course of disruptive political, economic or social crisis, it may be necessary for authorities (government, local municipalities, etc.) to implement appropriate intervention measures. However, experience from the recent pandemic also showed that the conditions for optimal, data-driven decisions are not yet available. One of the main limitations is the lack of sufficiently detailed information on critical goods and services (e.g. food, fuel, medical services) that could be used for early detection, as well as to define the ideal response strategy. The Dagmar¹ project develops data-driven tools to be used by authorities in such crisis situations.

There is already a legal basis for authorities to access such information for crisis prevention and management. On the European level, the Data Act enables public sector bodies to access and use data held by the private sector for specific public interest purposes. In Austria, a set of economic management laws (Lebensmittelbewirtschaftungsgesetz 1997 (LMBG), BGBl. Nr. 789/1996 idF. BGBl. I Nr. 113/2016, Energielenkungsgesetz (EnLG 2012) BGBl. I Nr. 41/2013 idF. BGBl. I Nr. 68/2022 and Versorgungssicherungsgesetz (VersssG 1992) BGBl. Nr. 380/1992 idF. BGBl. I Nr. 94/2016) enable authorities to implement counter-measures for specific critical supplies (e.g. food, energy) and access the corresponding data. However, the corresponding technical framework and implementation is missing.

To enable a data-based crisis management system, we need to make the data available and searchable, so that various applications, such as dashboards or question-answering systems, can be developed on top of it. However, most of the data in this scenario is not publicly accessible, but instead only available for certain actors (e.g. authorities) under certain conditions (e.g. when

¹<https://projekte.ffg.at/projekt/5120353>

a certain alert limit has been reached or during a crisis situation) – both of which may also change with time. Data spaces provide a basis for sharing data with well-defined policies, while OBDA provides a way for fast and efficient queries over data sources. However, we also need to be able to limit access to query results according to the dataset-specific usage right as specified by the corresponding policies.

2.1.2. Manufacturing

The UNDERPIN² project develops and deploys a data space for critical manufacturing sectors, where we explore two application areas, refineries and wind farms, for dynamic asset management as well as predictive and prescriptive maintenance.

For refineries, the aim is to improve the maintenance process and decision making to determine the best timing for preventive maintenance so as to minimize the downtime and the impact on the production capabilities. The wind farms use case is about optimizing the maintenance of wind farms, where Wind Turbine Generators (WTG) are deployed. WTG failure can have various reasons, with gearbox failures being one of the most common ones, and maintenance tasks as well as downtimes have high associated costs.

Data sharing along the value chain in such application areas is crucial due to the fragmented nature of data access, hindering the effective implementation of machine learning (ML) models. Each manufacturer may use different sensors, data formats, and communication protocols, resulting in fragmented data. With various stakeholders involved, access to all relevant data becomes a challenge. A holistic approach requires integrating data from multiple stages of the value chain, which may be managed by different stakeholders and systems. Integrating these diverse data sources in a beneficial way for all stakeholders requires harmonization efforts.

To address this, OBDA-based data consolidation (see Sec. 2.3.2) emerges as a viable solution. By employing standardized ontologies, stakeholders can efficiently and in a unified way access relevant data, streamlining the integration process and facilitating the utilization of ML models on a larger and more diverse data basis for improved decision-making and operational efficiency across the value chain. In the concrete use cases around predictive maintenance for wind turbines and refineries, developers of ML models could search for relevant datasets (e.g. a particular class of sensor data from the type of machines of interest) based on the ontologies, describe and integrate data models of machines and use such data for training, increasing the quality of their predictions.

2.2. Data Spaces

Data spaces provide a distributed architecture for cross-organisational data exchange while maintaining data sovereignty, that is, where the data owner retains control over the access and usage of their own data. In this way the data economy is supported twofold: providing technical and non-technical standards for sovereign data exchange and trust.

Different initiatives address different aspects of data space building blocks. Here, we focus on the technical features, but there are also on-going efforts to clarify the business, legal and

²<https://underpinproject.eu/>

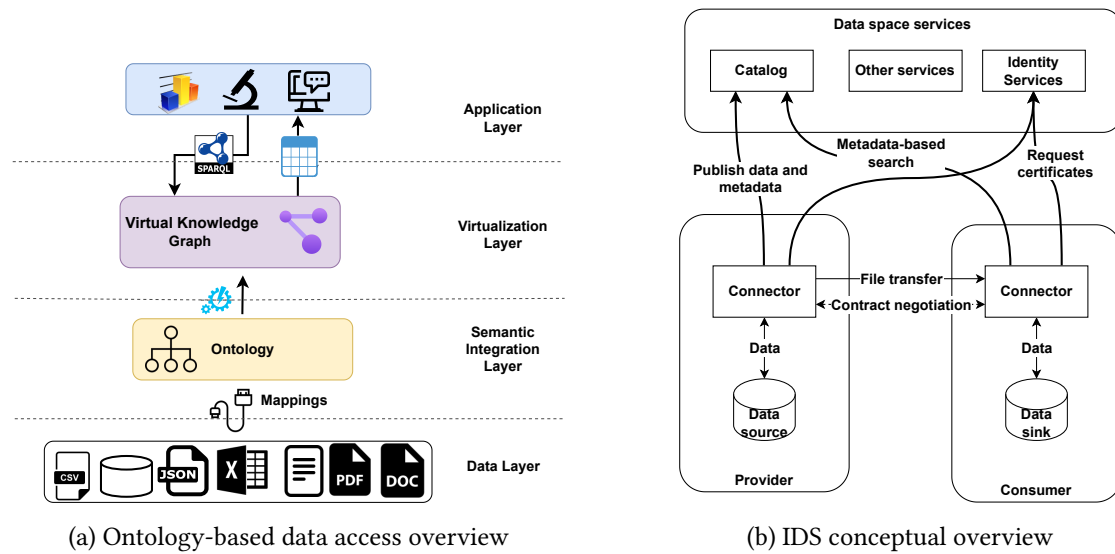


Figure 1: Ontology-based data access versus data spaces paradigms.

governance aspects of data spaces. The International Data Spaces Association (IDSA)³ is an initiative to define a standardized model and architecture for secure and trusted data exchange to drive the digital economy, as described in the International Data Spaces Reference Architecture Model (IDS-RAM)⁴. GAIA-X⁵ envisions a service architecture built on three pillars: compliance, federation (multiple actors cooperating based on shared rules) and data exchange. Gaia-X currently focuses on compliance, established on the basis of a decentralized trust framework, but also provides specifications for federations and data exchange.

There are also coordination initiatives integrating different frameworks. The Big Data Value Association (BDVA)⁶ is an industry-driven research and innovation organisation with a mission to develop an innovation ecosystem that enables the data-driven and AI-driven digital transformation of the economy and society in Europe. The Data Spaces Support Centre (DSSC)⁷, funded by the European Commission, supports the creation of data spaces with the aim of enabling data reuse within and across sectors to support the European economy and society. On the technical level, Simpl⁸ is an upcoming open source, smart and secure middleware platform that supports data access and interoperability among European data spaces, also funded by the European Commission.

Data semantics play an important part in data spaces to provide FAIR principles [4] for data sharing as Semantic Web standards provide a solid formal basis to define and publish vocabularies and ontologies. Moreover, shared and open standards for the semantic descriptions

³<https://internationaldataspaces.org/>

⁴<https://docs.internationaldataspaces.org/ids-ram-4/>

⁵<https://gaia-x.eu/>

⁶<https://bdva.eu/>

⁷<https://dssc.eu/>

⁸<https://digital-strategy.ec.europa.eu/en/policies/simpl>

of the metadata of data assets, and also for the data itself, support semantic interoperability[5]. Such descriptions make data in the data space *findable* via catalogs, automatically *accessible* via software components, *interoperable* by linking descriptions together and *reusable* based on standard descriptions. Furthermore, Semantic Web standards can also be used to describe the contracts and obligations for data sharing. In particular, the Open Digital Rights Language (ODRL) [6] provides a common basis to formulate contract requirements (policies), which can also be processed automatically and thereby provide a technical basis for automatic policy enforcement.

2.2.1. IDS Data Space Architecture

Our data space concept is built on the IDS architecture as depicted in Figure 1b. We chose IDS because of its widespread use in projects, like UNDERPIN, availability of implementations such as EDC⁹ and uptakes in practice e.g. mobility-dataspace¹⁰. We briefly describe the components of the IDS reference architecture model (IDS-RAM). The core building block of an IDS data space is the *Connector*. A *Connector* is a software component, which represents a participant of the data space, and which can provide and consume data under contractual obligations. In other words, a *Connector* is a gateway to secure data sharing in a data space. For security purposes, the IDS-RAM includes a *Certificate Authority (CA)* and a *Dynamic Attributes Provisioning Service (DAPS)* for dynamic access management, which manages dynamic access tokens. While *Connectors*, *CA* and *DAPS* are mandatory, the following components are optional based on specific needs of a data space:

- *Meta Data Broker* acts as a central metadata index-service, where connectors can publish information about data assets based on FAIR principles.
- *Vocabulary Hub* provides detailed (semantic) data models, which are used by the *Meta Data Broker* for metadata descriptions.
- *App Store* provides a platform for secure data apps which run in conjunction with a *Connector*.
- *Clearing House* provides a central logging service for clearing and billing as well as usage control.

2.2.2. IDS and Semantic Web Standards

The International Data Spaces Association publishes an information model (IDS-IM) [7] for data spaces, which is based on Semantic Web standards like RDF and OWL for interoperability, trust and usage control. Especially on the interoperability side, Semantic Web technologies provide flexibility for data modelling and integration of data and metadata alike.

The IDS-RAM includes components to enable interoperability for shared assets on the data and metadata level. Specifically, the *Meta Data Broker* acts as a metadata repository for published assets and enables the adherence to the FAIR principles based on RDF vocabularies. The *Vocabulary Hub* hosts (standard) RDF-based vocabularies used in the metadata descriptions

⁹<https://projects.eclipse.org/projects/technology.edc>

¹⁰<https://mobility-dataspace.eu/>

of the assets, where the data can be made available using standards like SPARQL [3] for easy integration. The vocabularies can also be used to describe the data itself by applying them to unstructured and structured data. For unstructured data using natural language, principles like semantic annotations can be applied. Vocabularies based on the Simple Knowledge Organisation System (SKOS) [8] provide concepts with multilingual labels which can be effectively used for semantic annotations of textual content. For structured data, there are different ways to map and transform towards an RDF-based representation, such as the RDF Extension for OpenRefine [9], for example. For relational data, W3C provides two recommendations, which are a direct mapping to relational data [10] and the mapping language R2RML [11].

2.3. Semantic Web Technologies for Semantic Interoperability

We describe below two existing approaches that rely upon Semantic Web technologies for improving semantic interoperability between datasets.

2.3.1. Vocabulary-based Data Access

When it comes to arbitrary data sources, a lightweight approach to support interoperability is by using published standard vocabularies to describe the metadata of each dataset. A well established W3C recommendation, which has seen wide adoption, is the Data Catalog Vocabulary DCAT [12]. DCAT is designed to facilitate interoperability between published data sets using a lightweight and generic approach to descriptions. DCAT is based on RDF and thereby can be easily extended and complemented with other vocabularies, expanding the descriptions of datasets based on FAIR principles. However, DCAT focuses on metadata only and does not provide ways to consolidate different vocabularies for descriptions, i.e. defining relations between different vocabularies, which can increase the level of interoperability in practice. An approach to achieve these two requirements for annotating both metadata and data and to interlink different vocabularies are taxonomic crossovers [13]. Taxonomic crossovers enable us to interlink concepts within different vocabularies or concept between different versions of the same vocabulary to achieve interoperability. Having established taxonomic crossovers for interoperability, we can automatically leverage them based on Semantic Web standards by evaluating them using SPARQL. When looking at the interoperability requirements for data spaces, where many participants can provide data assets and there is a high need for interoperability, such a solution is a major improvement for collaboration.

2.3.2. Ontology-based Data Access

The ontology-based data access (OBDA) paradigm enables access to a variety of disparate and heterogeneous data sources by semantically mapping information of each data source to the concepts and relations defined in a domain-specific ontology [2]. A key notion in OBDA is that of a *mapping* which consists of: (i) a *source* query, in the language of the data format, which extracts the relevant data values, and (ii) a *target* declaration which describes how the result of the query should be interpreted based on the domain ontology. An example of a mapping over a relational database is as follows:

```
@prefix sc: <http://www.dataspace-supplychain.com/> .  
@prefix owl: <http://www.w3.org/2002/07/owl#> .
```

```

source SELECT prodID, country, pname FROM product
        INNER JOIN exportban ON prodID
target sc:product/{prodID} a sc:Product ;
        sc:name "{pname}" ;
        sc:hasExportBan sc:country/{country} .
sc:country/{country} a sc:Exporter/{pname} .
sc:Exporter/{pname} owl:subclassOf sc:Exporter .

```

In this mapping, the source declaration is an SQL query that looks up the product id, name and country where this product is banned from being exported. The target declaration creates the following RDFS assertions: class `sc:Product` is instantiated using product ids, an individual for each country exporter is created as an instance of a new product exporter class which in turn is a subclass of `sc:Exporter`.

A result is a map of the form $(prodID, country, pname) \mapsto (0142, China, Germanium)$ yielding the following RDFS assertions (in Turtle syntax):

```

@prefix sc: <http://www.dataspace-supplychain.com/> .
@prefix owl: <http://www.w3.org/2002/07/owl#>
sc:product/0142 a sc:Product ;
        sc:name "Germanium" ;
        sc:hasExportBan sc:country/China .
sc:country/China a sc:Exporter/Germanium .
sc:Exporter/Germanium owl:subclassOf sc:Exporter .

```

Another important element in OBDA is the *virtualization* of data, meaning that the actual transformation of the data into RDF(S) and storage of the knowledge graph is not materialized. In this approach, the ontology and mappings for each data source, also denoted as OBDA specification, expose the underlying data as a virtual set of RDF(S) assertions, making it accessible at query time using SPARQL. This mechanism is realized by transforming each SPARQL query using the OBDA specification into a set of format-specific queries over each data source, then aggregating the answers.

3. Information Access in Data Spaces via OBDA

In this section we describe how OBDA can be realized within the IDS model to enable instant access to relevant information using Semantic Web technologies, while covering the requirements imposed by the data space framework. The envisioned conceptual model of a data space, that supports efficient search and retrieval of relevant information for a data consumer while preserving the policies of data access instated by automatic contractual agreements with data providers, is presented in Figure 2. We consider three required phases: setting up a data space for data exchange, publishing data in the data space, and finally searching the available data.

3.1. Setup for Building a Data Space

First we need to provide the basis for data sharing, where participants are willing and able to provide and consume data, while respecting data sovereignty. The rules and building blocks for

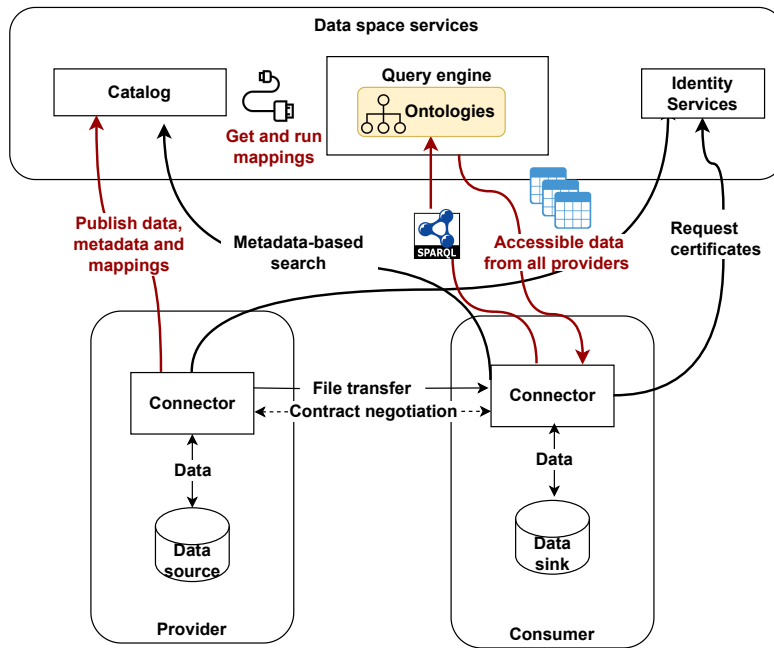


Figure 2: Conceptual view of the data spaces using OBDA for accessing and sharing data

both business/organisational and technical aspects need to be defined and implemented and participants need to be on-boarded to the system in accordance with these rules.

For participants, verifiable descriptions with a set of mandatory attributes could be provided by the participant and verified as part of the on-boarding process. The data model for such participant descriptions needs to be stored at a reliable location (e.g. decentralised storage or a trusted central authority) and made publicly accessible .

For assets, extendible base models may be defined and managed by a component provided for all data space participants (semantic hub in EDC). These models may also include links to the relevant domain specific ontologies.

Each data space is related to a specific domain, for which an ontology can be developed or curated from existing vocabularies. The ontology should define all the key concepts and relations to describe which information is relevant for the majority of the stakeholders in the data space thus enabling a common understanding of the data that is being exchanged. In the context of the Dagmar project, most of the stakeholders are involved in designing the ontology. These efforts will enable a common understanding of the data being exchanged, therefore this task is done in parallel with the task of designing the data space.

3.2. Publishing Data

In our approach a data provider is advised to publish the metadata and the ontology mappings alongside the data itself. A main benefit about this new approach is the modularization of the data, as access can be moved to the level of particular information within the dataset, but

the data provider still has control over the mapping definitions and thus can restrict access to sensitive information.

The creation of the mappings can be challenging for non-expert users, therefore several techniques for automatic generation have been proposed in the literature, such as using machine learning solutions [14] or editing tools for manual crafting of the mappings [15]. Within the dataspace framework, data providers should receive technical support and services to create the mappings, therefore this aspect must be taken into account in designing the dataspace.

The access and usage policies are established and clearly defined by the data owner. If some participants should have special data access privileges, their participant descriptions (credentials) need to include the information to enable granting these; as described by the policy specific to the data asset. During the querying time, the credentials and policies are then taken into account when retrieving answers.

For example, the access levels can be defined as follows:

- **Allow**: enabled for all users who have valid credentials to access the information.
- **Restricted**: enabled for some users that have valid credentials and satisfy the corresponding policies.
- **Disallowed**: disallowed for all users.

On a technical level, the levels can be defined in the form of an ODLR policy, with different rules specified for different assignees (recipients of the rule, i.e. the user consuming the data).

Then the mapping language can include such access schema to the level of the target declarations. For example, if we want to restrict all information related to export bans, we can update the previous mapping:

```
@prefix sc: <http://www.dataspace-supplychain.com/> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
source SELECT prodID, country, pname FROM product
INNER JOIN exportban ON prodID
target sc:product/{prodID} a sc:Product ;
                                sc:name "{pname}" ;
                                sc:hasExportBan sc:country/{country} . @restricted
sc:country/{country} a sc:Exporter/{pname} . @restricted
sc:Exporter/{pname} owl:subclassOf sc:Exporter . @restricted
```

This would generate the following restricted RDF graph to users that have valid credentials but do not satisfy the requirements described by the dataset's policy:

```
@prefix sc: <http://www.dataspace-supplychain.com/> .
@prefix owl: <http://www.w3.org/2002/07/owl#>
sc:product/0142 a sc:Product ;
                sc:name "Germanium" .
```

Note that the access level can be always updated in the mappings, without the need to change anything else in the approach. In the case that the ontology mappings are not provided, then the standard search and exchange mechanism remain in place.

3.3. Searching in the Data Space

Next, we need to search mappings of available data sources in the data space, but without revealing any sensitive information (policy-protected OBDA). We first describe two options for

the general search process and then describe possible queries and their evaluation in detail.

Search process We envision two different solutions: a one-step constrained search, or a two-step approach of search on accessible data; both conceptually interoperable with the mutual contracting method defined in the IDS dataspace protocol.

In the one-step constrained search, the search itself would be a special case of data access consisting of the following steps:

1. Consumer formulates search query and sends it to the query engine.
2. Query engine requests results for consumer from each provider.
3. Providers evaluate the query, with constraints applied from the policies combined with the consumer's identity.
4. Query engine combines results and returns them to the consumer.

In the two-step approach of search on accessible data, we propose a preparation phase and a query phase. In the preparation phase, searchable assets in the data space are assembled for a given consumer by evaluating the policies of each asset to determine if they are searchable for that consumer. Then, when the consumer formulates a query, an unconstrained search can be performed. The query step in this method is simple and quick, since policies are enforced in the preparation step. However, since the preparation step is costly and only performed on demand, this is only suitable if (1) consumers are known in advance and (2) the set of available data assets in the data space is stable. This is the case for our supply chain resilience use case, but not for our manufacturing example.

Search queries In either approach, the data consumer can access the data catalog and then use the standard metadata-based search to find relevant datasets and their providers. Additionally, the consumer can also pose SPARQL queries based on the ontology such as:

```
SELECT *  
WHERE {  
  ?product a sc:Product ;  
    sc:name ?name ;  
    sc:hasExportBan ?country  
}
```

Based on a query, that encodes the information needs, the data provider can either: (a) **Verify access:** Request to verify if the query is allowed, given the participant's credentials, and if there are any matches on specific datasets. (b) **Get answers:** Request to construct answers to the query, with the option of selecting the datasets of interest.

Considering the above query and the previous mapping example that restricts the access to exports bans, the answer to request (a) would be "no" if no special privileges are in place, and "yes" otherwise. In the case of "yes", the user can use service (b) and gets the following answer:

<i>?product</i>	<i>?name</i>	<i>?country</i>
sc:product/0142	"Germanium"	sc:country/China

4. Related Work

Auer et al [5] discusses the potential of using Semantic Web technologies for achieving semantic interoperability in data spaces. Similarly, Theissen-Lipp et al. [16] also point out the potential use of ontologies to mitigate the access to data within the data space. However, no concrete conceptual model is proposed and the problem of mapping the data to an ontology in a data space is not addressed.

In Boukhers et al.[14], the authors propose the use of machine learning techniques for automatic meta-data extraction and ontology alignment as well as for mappings generation. Such techniques are still applicable in our framework to ease the creation of the mappings and to improve searchability of datasets, however our focus is on querying the datasets and how can OBDA paradigm be used within a data space. In Langer et al.[17], the authors propose the use of ontologies to mediate the access to the datasets, however without the consideration of mappings and access restriction.

Regarding OBDA approaches that can be applied for data spaces, existing approaches that support access control have been proposed [18, 19], however access rights and control is modeled in the ontology or the access restrictions are placed upon the properties in the ontology. Cima et al.[20] introduced the notion of policy-protected OBDA (PPOBDA), where an OBDA specification (consisting of the domain ontology, schema and mapping) is extended by a set of policy constraints. The authors describe a method to reduce PPOBDA specifications to OBDA specifications that keep the same domain ontology and schema, but incorporate policy constraints into the mapping. They also conduct experiments to show runtimes on a set of SPARQL queries in this setting. Our approach is related to PPOBDA and their solutions can still be applied in our conceptual model tailored for data spaces.

5. Discussion and Conclusions

In this paper we presented a conceptual model for adapting ontology-based data access paradigm to enhance searchability and semantic interoperability in data spaces.

We described two motivating examples for which OBDA functionalities in the data space is highly beneficial. In our novel conceptualization of a data space, we propose to publish the data assets alongside metadata and additionally with the mappings to a domain specific ontology that enable searching the data in the entire data space. Due to the shared conceptualization encoded in a common ontology, information that comes from multiple sources can be retrieved using the same query. This mechanism is enabled by mapping each data asset to the data space ontology. To restrict access to some particular information, we proposed to add access restriction labels as part of the data asset policy in the mapping declaration. We also described and exemplified the searching and data access mechanism in our novel data space framework. As outlined below, we address some of the existing open points in our approach and potential challenges.

A first observation is that our paper focuses on an architecture based on the IDS RAM and in particular on the notion of connectors handling all operations on behalf of a participant. Our concepts are agnostic to the exact specification of the connector, but would have to be slightly adapted for a connector-less design, such as the blockchain-based system of Pontus-

X [21]. In such a case, the queries could be initiated directly by the participant, e.g. via a central management platform, which would send the request, together with the participant's credential, and trigger the query. Such systems also normally include a non-connector-based policy enforcement engine, which could be extended to handle policy enforcement for OBDA queries.

A second observation is regarding the ontology creation and maintenance procedure. The design of the ontology has to be discussed among all relevant stakeholders, however if there exists some governing entity, then, in principle, it can take the responsibility to design and maintain the ontology.

A third observation is about feasibility in practice, namely checking access control for each query which can be problematic for the query engine system. However due to the static nature of the credentials of each consumer to each dataset and the fact that the mappings are not frequently updated, the mappings-based access credentials can be computed in advance and efficiently stored and used at query time (see two-step approach in subsection 3.3).

Last but not least, ontology reasoning has to be taken into account when accessing and computing the answers to queries. For instance, if a property has restricted access in a mapping and in a query a sub-property is being used, then the query should not have access to the data. For this challenge query evaluation techniques such as the one proposed in [20] can be used.

Acknowledgments

This work was funded by the Austrian security research program KIRAS of the Federal Ministry of Finance (BMF) through the DAGMAR project (grant No. 52224305), Austrian Research Promotion Agency (FFG) under grant No. FO999913202 *UNDERPIN* as well as by the European Commission under contract No. 101123179 *UNDERPIN*.

References

- [1] B. Otto, M. Hompel, S. Wrobel, *Designing Data Spaces: The Ecosystem Approach to Competitive Advantage*, Springer International Publishing, 2022. URL: <https://books.google.at/books?id=gfbWzgEACAAJ>.
- [2] A. Poggi, D. Lembo, D. Calvanese, G. D. Giacomo, M. Lenzerini, R. Rosati, Linking data to ontologies, *J. Data Semant.* 10 (2008) 133–173. URL: <https://api.semanticscholar.org/CorpusID:1325494>.
- [3] E. Prud'hommeaux, S. Harris, A. Seaborne, *SPARQL 1.1 Query Language*, Technical Report, W3C, 2013. URL: <http://www.w3.org/TR/sparql11-query>.
- [4] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, et al., The fair guiding principles for scientific data management and stewardship, *Scientific data* 3 (2016) 1–9.
- [5] S. Auer, Semantic integration and interoperability, in: *Designing Data Spaces*, Springer, 2022, pp. 195–210.
- [6] R. Ianella, *Open digital rights language (odrl)*, *Open Content Licensing: Cultivating the Creative Commons* (2007).

- [7] C. Lange, J. Langkau, S. Bader, The ids information model: a semantic vocabulary for sovereign data exchange, *Designing data spaces* (2022) 111.
- [8] A. Miles, S. Bechhofer, SKOS Simple Knowledge Organization System Reference, Working Draft, W3C, 2008. URL: <http://www.w3.org/TR/skos-reference>.
- [9] R. Verborgh, M. De Wilde, *Using openrefine*, Packt Publishing Ltd, 2013.
- [10] M. Arenas, A. Bertails, E. Prud'hommeaux, J. Sequeda, et al., A direct mapping of relational data to rdf, *W3C recommendation 27* (2012) 1–11.
- [11] S. Das, R2rml: Rdb to rdf mapping language, <http://www.w3.org/TR/r2rml/> (2011).
- [12] R. Albertoni, D. Browning, S. Cox, A. N. Gonzalez-Beltran, A. Perego, P. Winstanley, The w3c data catalog vocabulary, version 2: Rationale, design principles, and uptake, 2023. [arXiv:2303.08883](https://arxiv.org/abs/2303.08883).
- [13] A. Ahmeti, J.-K. Schakel, R. David, A. Revenko, Towards preserving biodiversity using nature first knowledge graph with crossovers (2023).
- [14] Z. Boukhers, C. Lange, O. Beyan, Enhancing data space semantic interoperability through machine learning: a visionary perspective, in: *Companion Proceedings of the ACM Web Conference 2023, WWW '23 Companion*, Association for Computing Machinery, New York, NY, USA, 2023, p. 1462–1467. URL: <https://doi.org/10.1145/3543873.3587658>. doi:10.1145/3543873.3587658.
- [15] A. Paulus, A. Pomp, T. Meisen, The plasma framework: Laying the path to domain-specific semantics in dataspaces, in: *Companion Proceedings of the ACM Web Conference 2023, WWW '23 Companion*, Association for Computing Machinery, New York, NY, USA, 2023, p. 1474–1479. URL: <https://doi.org/10.1145/3543873.3587662>. doi:10.1145/3543873.3587662.
- [16] 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. URL: <https://doi.org/10.1145/3543873.3587689>. doi:10.1145/3543873.3587689.
- [17] T. Langer, A. Pomp, T. Meisen, Towards a data space for interoperability of analytic provenance, in: *Companion Proceedings of the ACM Web Conference 2023, WWW '23 Companion*, Association for Computing Machinery, New York, NY, USA, 2023, p. 1502–1503. URL: <https://doi.org/10.1145/3543873.3587686>. doi:10.1145/3543873.3587686.
- [18] C. Choi, J. Choi, P. Kim, Ontology-based access control model for security policy reasoning in cloud computing, *J. Supercomput.* 67 (2014) 711–722.
- [19] C. Brewster, B. Nouwt, S. Raaijmakers, J. Verhoosel, Ontology-based access control for FAIR data, *Data Intell.* 2 (2020) 66–77.
- [20] G. Cima, D. Lembo, L. Marconi, R. Rosati, D. F. Savo, Controlled query evaluation in ontology-based data access, in: *ISWC (1)*, volume 12506 of *Lecture Notes in Computer Science*, Springer, 2020, pp. 128–146.
- [21] deltaDAO AG., *Pontus-X Documentation*, 2024. URL: <https://docs.pontus-x.eu/>.