2025 European Conference on Computing in Construction
CIB W78 Conference on IT in Construction
Porto, Portugal
July 14-17, 2025

VOCABULARY HUB FOR SEMANTIC DATA EXCHANGE AND INTEROPERABILITY IN AN AECO DOMAIN-SPECIFIC DATA SPACE

Avinash Ramachandruni, Ravi Sharma, Ekaterina Petrova, and Pieter Pauwels
Eindhoven University of Technology, Eindhoven, The Netherlands

Abstract

The digitalisation of the built environment faces challenges such as interoperability, standardisation, integration of legacy systems, and secure data exchange. The International Data Spaces (IDS) framework, used successfully in logistics and energy, enables organisations to create value through data. This paper examines how the Vocabulary Hub, a core IDS component, addresses these challenges in the Architecture, Engineering, Construction, and Operation (AECO) sector. It outlines the setup of a domain-specific vocabulary hub and analyses its role in semantic data exchange, suggesting that IDS is a promising approach for data integration in federated digital twin systems and structured data sharing in AECO.

Introduction

In the AECO industry, efficient information exchange between stakeholders and systems is crucial to project success. The sector has evolved from manual data sharing to automated data flows, yet the challenge of breaking system silos persists. Organisations are losing the significant advantages of federated systems, which enable seamless data exchange, improve operational efficiency, and potentially reduce costs (Taboada-Orozco et al., 2024).

Digital technologies such as Building Information Modelling (BIM), the Internet of Things (IoT), and Cloud Computing have revolutionised modern buildings, making them smarter and more data driven. However, the increasing complexity of building systems requires more efficient data-exchange methods. Traditional approaches, such as middleware, APIs and data warehousing are inherently centralised, limiting scalability and ease of integration. The adoption of Digital Twins (DT) in the construction sector calls for the development of advanced frameworks that not only enable data exchange but also facilitate the integration of multiple data sources (Čustović et al., 2023).

The selection of a data exchange framework depends on the business needs of an organisation and also on the size of the project. Despite this contrast, there are few common challenges that we observe in the industry. These challenges are related to data standardisation and consistency, interoperability, security, scalability, integration complexity, and real-time integration. Tang et al. (2019); Afsari et al. (2017) suggested cloud-based approach as a future direction for data management, particularly for cases requiring real-time synchronisation and analytics.

The IDS framework, designed for data sovereignty, is gaining attention in the architectural and construction sectors as a decentralised data exchange solution. Building a data space often involves linking multiple domains (e.g. energy prediction, weather data) that are relevant for a use case. The IDS framework is also suitable for a distributed data federation system such as the federated building or the urban digital twin. The different components of an IDS are shown in Figure 1. The central component of a data space is the connectors. The interaction between the components occurs through these connectors. The vocabulary hub is the place to store and access the vocabularies. The metadata broker enables data providers to store the self-descriptions of their data assets. Hence, for a semantic data exchange, these three components are primary supported by other data space components.

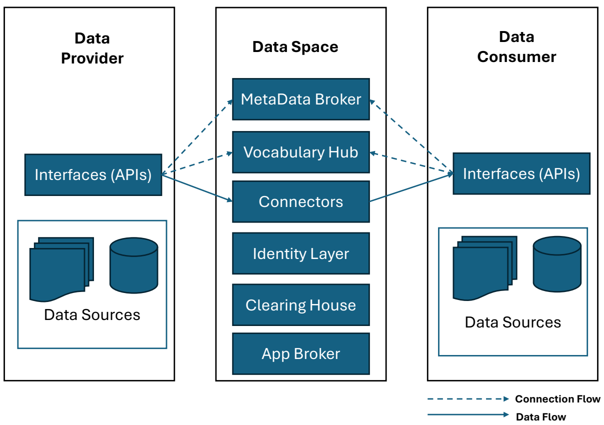

Figure 1: IDS Components Diagram as per IDS-RAM4.

This paper addresses the use of the IDS as a decentralised data exchange framework. The study also analyses the role of a vocabulary hub as specified in the Reference Architecture Model (RAM) (IDSA (2025)) for data standardisation, quality, and semantic interoperability, and defines the steps to create a domain-specific vocabulary hub from AECO.

Literature Review

Building Information Modeling (BIM) has been a key development in sharing information between stakeholders in the construction project. However, the use of proprietary software limits the interoperability of designs and project information. To achieve interoperability for optimizing performance (Farghaly et al., 2023), standards such as Industry Foundation Classes (IFC) (Costin et al., 2022) are developed. These standards, often called an ontology or a vocabulary, enable a shared understanding of concepts without requiring a unified data model, which is impractical for large-scale complex systems and the semantic web. For the scope of this paper, we review the state-of-the-art related to data exchange frameworks, ontologies in AEC and semantic interoperability in the data spaces. In AECO, many studies have been done that address the interoperability issue, particularly for the usage of BIM.

Most existing studies first aim for semantic data exchange rather than full interoperability (Pauwels et al., 2017). The same study also presents that semantic web technologies will aid in creating common data models such as Common Data Environments (CDE). The ever growing volumes of organizational data suggest that we move towards decentralized data exchange. Table 1 details the various data exchange frameworks that are widely used and how they measure against each of the challenges discussed earlier. The assessment is based on the chapter on the measurement of architectural characteristics in Richards (2022)). This comparative analysis helps motivate the selection of architectural styles suitable for the development of a domain-specific data space. Although broader characteristics such as scalability, security, and integration complexity are addressed primarily by the overarching IDS framework, this paper specifically examines how the vocabulary hub contributes to enhancing other critical aspects, particularly data consistency, interoperability, and flexibility, to achieve efficient semantic data exchange. Many past studies are also towards development of ontologies that are either new or for mapping data schemas (Farghaly et al., 2023) across domains (e.g. CityGML). Modularization, merging, and mapping of the ontology is also considered an important aspect of data exchange. Smaller ontologies help data development teams work more efficiently and reduce complexity. However, this approach can discourage the reuse of existing ontologies, as creating a new one may seem easier than adapting an existing model. The merging and mapping steps help to extend the existing ontology or to create semantic bridges between ontologies of various domains (Farghaly et al., 2023).

Interoperability goes beyond system integration or linking of data sources (Pauwels et al., 2017) by using ontologies to enrich the data, make it consistent and machine readable, and enable automatic reasoning over the data assets or data sources. Steinbuss et al. (2024) highlights semantic interoperability as a critical component of the New European Interoperability Framework (EIF), which defines four layers of interoperability: technical, semantic, organisational, and legal. Given that data spaces consist of heterogeneous resources, technical interoperability (standard data exchange formats and protocols) becomes a prerequisite for semantic interoperability. Moreover, semantic interoperability in a domain-specific data space is considered essential for federated systems (Čustović et al., 2023; IDSA-Rulebook, 2025), including Building and Urban DT.

IDS for a Decentralized Data Exchange and Integration

In a data-driven smart built environment, user applications expect low latency when visualising the results of complex queries. To enable real-time data delivery, a common approach is to process raw data at the source - similar to edge computing in IoT - and share only aggregated information with a central location (in centralised systems) or directly with consumers (in decentralised systems). Although this reduces network load, it also creates challenges in combining data from multiple sources to obtain comprehensive insights.

For example, in a smart office building, real-time occupancy data from IoT sensors can be integrated with energy consumption data from the Building Management System (BMS) and weather forecast data from an external API to dynamically optimise HVAC settings. Although edge devices can process local data for immediate responses, achieving holistic building intelligence requires

Table 1: Comparision of performance of exchange frameworks against challenges for data management

Challenge	Point-to-Point (API)	Middleware	Centralized Platforms	Decentralized Frameworks (IDS, Blockchain)
Ease of Integration (Complexity)	Low	Medium	Medium	Medium
Scalability	Low	High	Medium	Very High
Data Consistency	Medium	High	Medium	High
Security and Privacy	Medium	High	Low	High
Interoperability	High	Medium	Low	High
Flexibility	Low	Medium	Low	Very High

a centralised or federated data exchange framework to integrate multiple data streams across floors, departments, or entire facilities.

The growing adoption of DT in both the AEC and Facility Management (FM) industries has further driven the shift toward federated data sources. Organisations advocating for federated architectures also recognise the need to eliminate centralised control, which often becomes a bottleneck when exchanging large volumes of data. IDS framework plays a key role in addressing these challenges by enabling decentralised data exchange and integration while ensuring data sovereignty and trust through standardised protocols and a vocabulary hub.

Among the various components of IDS, the Vocabulary Hub is crucial for enabling semantic data exchange and integration. The following sections explore its role in the data exchange process and outline the steps required to build an AECO domain-specific vocabulary hub.

The Vocabulary Hub of a Data Space

A vocabulary hub is an IDS component that serves as a repository for vocabularies, including ontologies, reference data models, and taxonomies. In a study on data space ecosystems, Otto et al. (2022) proposed three evolutionary stages: closed, open, and federated. The federated stage includes both multilateral core participants and intermediaries (also called federators). This study focusses on the federated stage, as it is particularly relevant to develop a digital twin. According to IDS-RAM 4, the vocabulary hub functions as an intermediary that connects two data space participants, typically a data provider and a data consumer.

From an interoperability perspective, a federation of ecosystems requires open standards between domains and between domains, ontology mapping and translations, and unique identifiers between domains (Otto et al., 2022). These aspects form the core objectives of a vocabulary hub. Within a data space, a vocabulary hub generally follows a centralised architecture for vocabulary management. However, this study does not address the management of the vocabularies used to describe the data space itself, its components, or interoperability between multiple data spaces. This aspect is particularly relevant at the federation-of-ecosystems level or when a resource belongs to more than one data space. The existing literature explores decentralised and federated approaches to the implementation of a vocabulary hub, but such implementations are beyond the scope of this study.

In general, the purpose of a vocabulary hub is not limited to storage but also includes functions such as creation, selection, editing, reading, and search capabilities through collaborative development. This functionality is aligned with the requirements of the information layer in IDS-RAM. However, in the AECO industry, most ontologies are static and undergo minimal updates over their lifetime. This characteristic limits the need for frequent modifications within the vocabulary hub, making its role in the AECO context slightly different from the broader IDS-RAM specifications. Additionally, a vocabulary hub can be developed into a fully functional component with APIs, web interfaces, and user management, using the data space base connector system. When designing a domain-specific data space for the AECO/FM industry, a read-only vocabulary hub that supports searching and retrieving representations from multiple published ontologies is often the most practical approach.

Beyond this, the vocabulary hub offers distinct capabilities compared to existing solutions such as GraphDB or the buildingSMART Data Dictionary (bSDD). Like bSDD, the vocabulary hub can act as a place to publish and access ontologies, but it is not tied to a central organisation and can be implemented in a technically correct and domain-specific way. It also goes beyond just publishing vocabularies — it supports dynamic binding of semantic descriptions to data assets, advanced metadata annotation, and ontology alignment between multiple organisations or data providers.

GraphDB, while useful as a scalable RDF triple store with SPARQL support, is a general-purpose database. It does not include ontology lifecycle management and does not support role-based access or semantic mediation across domains. Ontologies are typically not managed directly in triple stores, and the vocabulary hub offers more control over how ontologies are curated, aligned, and shared. However, components such as GraphDB can still be integrated within a vocabulary hub architecture to improve querying and reasoning, as shown in David et al. (2024).

Building a Domain-specific Vocabulary Hub

One of the early implementations of an open source IDS vocabulary hub was developed under the PLATOON H2020 EU project (Nagel et al., 2021). This project adopted a containerised architecture, in which each microservice was responsible for accessing an API endpoint to facilitate searching and querying vocabularies. The use of microservices and API-driven architecture enabled scalability, modular deployment, and interoperability across different platforms. The system architecture in PLATOON aligns with the service-oriented software architecture for federated systems as proposed by Chamari et al. (2023). This architecture emphasises loose coupling between components, making it well-suited for distributed and federated data spaces, where different entities manage their own vocabularies while still enabling seamless data exchange.

Another notable vocabulary provider is the International Data Spaces Association (IDSA), which has developed its own vocabulary service to support the semantic alignment of data models within trusted data spaces. The IDSA approach enables organisations to register, discover, and map their domain-specific vocabularies while maintaining compliance with data sovereignty principles.

The following are the key steps in building a vocabulary hub for a domain-specific data space. These steps assume that relevant domain-specific ontologies exist in RDF/Turtle formats:

Step 1. *Design the Vocabulary Hub Architecture*

A monolithic architecture might be suitable for a small data space. However, for larger federated data spaces, it is recommended to use a microservice architecture, as implemented in the PLATOON project, to ensure flexibility and scalability. Such an architecture also allows one to build native cloud applications over existing infrastructure and shared services such as orchestration and identity management. Hence, based on the conceptual framework of the vocabulary hub, we propose a layered system architecture as shown in Figure 2. Although the vocabulary hub can function with non-RDF vocabularies (such as CityGML, JSON/XML schema) , this architecture emphasises RDF-based ontologies as the preferred standard to enable semantic querying and alignment across the data space.

Step 2. *Implementation of Services*

The services layer handles tasks such as validation and vocabulary management by accessing vocabularies through their URIs rather than storing them directly. Metadata for these vocabularies and their URIs is maintained separately within a metadata broker for efficient management.

Step 3. *Publish Existing Ontologies*

Since the vocabulary hub supports RDF-based ontologies, it is straightforward to load them into a triple store (such as an RDF store or GraphDB). The IDS vocabulary can be stored alongside domain-specific ontologies (e.g., SOSA, BOT, BRICK, and SAREF). SHACL (a constraint language for validating RDF data) provides constraint definitions that can be applied to RDF data graphs, enabling the validation of both instance data and ontology structures published within the data space.

Step 4. *Publish Self-Descriptions in a Metadata Broker*

If a metadata broker (for example, IDS Metadata Broker) is available, register the metadata of the domain-specific ontology (for example, the Data Catalog (DCAT) description, SHACL constraints). If not, the metadata must be maintained manually as an RDF dataset and exposed via SPARQL endpoints. Publishing to a metadata broker is crucial for data discovery and ensuring interoperability.

These steps ensure that the vocabulary hub is configured properly and ready for use by the participants. Once the vocabulary hub is set up, querying via SPARQL endpoints is a straightforward approach.

Role of Vocabulary Hub for Semantic Data Exchange and Interoperability

IDS-RAM serves as a guiding framework to understand how data exchange occurs between providers and consumers while ensuring semantic interoperability. To examine the role of the vocabulary hub in standardisation, we analyse the RAM process layer, including the self-description of data assets and the IDS Information Model (IDM).

Implementing IDM for a Domain-specific Data Space

In theory, a vocabulary hub is not strictly required for data exchange. If two participants within a data space agree to share data using standardised JSON or XML schemas, the IDS framework can support this exchange. However, while such data sharing ensures syntactic compatibility, it lacks contextual meaning and provides limited value to consumers. Similarly, a vocabulary hub that merely stores syntactic schemas (e.g., XML Schema or JSON Schema) without supporting ontological structures is also possible, though it limits semantic interoperability. In this case, the exchanged data would ensure structural consistency but would still lack semantic understanding. Without a well-defined ontology to establish relationships and meaning, the data remains disconnected, limiting its usability for intelligent applications and cross-domain integration.

IDS-RAM stipulates that communication between participants must occur through connectors using the standard IDS protocol. Additionally, the Information Model defines a domain-agnostic language, known as the IDS Vocabulary, to represent core data space concepts, including key components such as connectors and self-descriptions. While the IDS vocabulary primarily focusses on representing the core components and interactions within a data space, such as participants, connectors, resources, and contracts, not on domain-specific data exchange., the DCAT ontology is primarily used for metadata management (in metadata broker). Both the IDS Vocabulary and DCAT are built on RDF-based OWL schemas, ensuring that they meet the semantic interoperability requirements of the system.

IDS-RAM recommends using RDF-based ontologies to enable better querying capabilities at runtime. However, for ontologies based on other languages (XSD, JSON-schema, EXPRESS), three possible approaches can be used to achieve standardisation in interoperability:

1. Convert semantic data to RDF/OWL compatible formats such as ifcOWL.
2. Use alignment or mapping ontologies to link domain-specific classes to the IDS vocabulary.
3. Incorporate non-RDF/OWL vocabularies into the vocabulary hub for describing related data assets while also providing public interfaces (preferably cloud-based) to facilitate data exchange within the data space. Although this approach offers a lower level of semantic interoperability, it highlights the advantage of supporting multiple vocabularies based on different schemas within the vocabulary hub.

Regardless of the chosen approach, the vocabulary hub should enable seamless data exchange. In the AECO domain, most existing ontologies are RDF-based, with notable exceptions like IFC and CityGML. Even IFC is also available in the OWL-based format (ifcOWL), which makes it compatible for use in the vocabulary hub, alignment may still be necessary for integration with other ontologies like BOT or SAREF. Furthermore, the vocabulary hub supports ontology mapping for domain-specific schemas that were not originally developed using OWL/RDF. Figure 3 presents a detailed hierarchical model of a vocabulary hub for an AECO data space:

- The foundation layer consists of formalisms and modelling languages such as RDF and GML, which serve as the basis for defining schemas.
- Based on these formalisms, the next layer includes schemas/ontology definition languages such as OWL, XSD, and EXPRESS.
- Above this, the data-space related vocabulary layer contains essential conceptual models and standards relevant to data exchange within the data space.
- Finally, the top layer is dedicated to domain-specific vocabularies, ensuring adaptability to different industries and specialised use cases.

In the context of a minimum viable data space, a vocabulary hub, if implemented, should include at least the first three layers to establish a foundational semantic framework for interoperability.

Self-Description of a Data Source and Relationship with Metadata Broker

A data provider must create a self-description (in JSON-LD, RDF, or TTL format) for any data asset within a data space before making it available for use by a data consumer (Steinbuss et al., 2024). IDS-RAM elaborates on the process of publishing, creating, and using vocabularies in two phases: design and runtime. During the design phase, publishing self-descriptions can be based on existing ontologies already in use within the data space or on newly created ontologies. If a new ontology is used, the data provider can publish it in the vocabulary hub. However, even existing ontologies must be published if the data asset is the first of its kind in the data space.

IDS-RAM allows the data provider to store multiple self-descriptions in another IDS component called the Metadata Broker (see Figure 1). The framework also establishes a clear distinction between the metadata broker and the vocabulary hub, defining their respective roles within a data space. A sample self-description for a temperature sensor is shown in Listing 1 and stored in the Metadata Broker.

Listing 1: The sample TTL mentioned below illustrates the Self-Description process for a sensor data asset.

```
@prefix ids: <https://w3id.org/idsa/core/> .
@prefix sosa: <http://www.w3.org/ns/sosa/> .
@prefix ifc: <http://ifcowl.openbimstandards.org/IFC2X3_Final#> .
@prefix sbm: <http://smartbuildingmanagement.org/> .

    <!-- Metadata related to IDS -->
  # IDS Connector representing the data provider
    sbm:connector123 a ids:Connector ;
        ids:title "Smart_Building_Connector" ;
        ids:description "Connector_for_smart_building_sensors" ;
        ids:maintainer sbm:SmartBuildingCorp ;
        ids:curator sbm:SmartBuildingCorp ;
        ids:securityProfile ids:BASE_SECURITY_PROFILE ;
        ids:catalog sbm:catalog123 .

    # Participant (data provider) information
    sbm:SmartBuildingCorp a ids:Participant ;
        ids:title "Smart_Building_Corp" ;
        ids:memberParticipant sbm:SmartBuildingCorp .

    # Catalog of offered resources
    sbm:catalog123 a ids:ResourceCatalog ;
        ids:offeredResource sbm:resource456 .

    # Resource representing the temperature sensor data
    sbm:resource456 a ids:Resource ;
        ids:title "Room_205_Temperature_Sensor_Data" ;
        ids:description "Room_205_sensor_readings" ;
        ids:representation sbm:representation789 ;
        ids:contractOffer sbm:contract101112 .

    # Representation of the data
    sbm:representation789 a ids:Representation ;
        ids:mediaType "application/json" ;
```

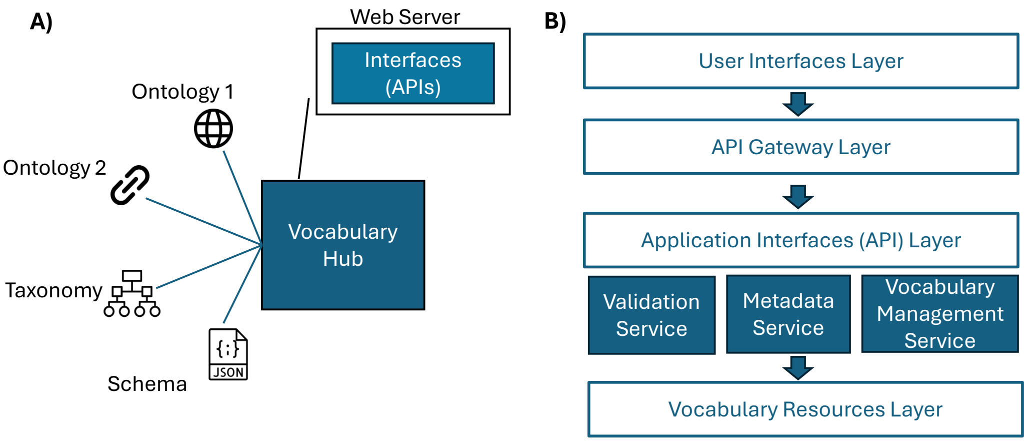

Figure 2: A) Conceptual Framework of Vocabulary Hub B) Proposed System Architecture for an AECO Domain Specific Vocabulary Hub.

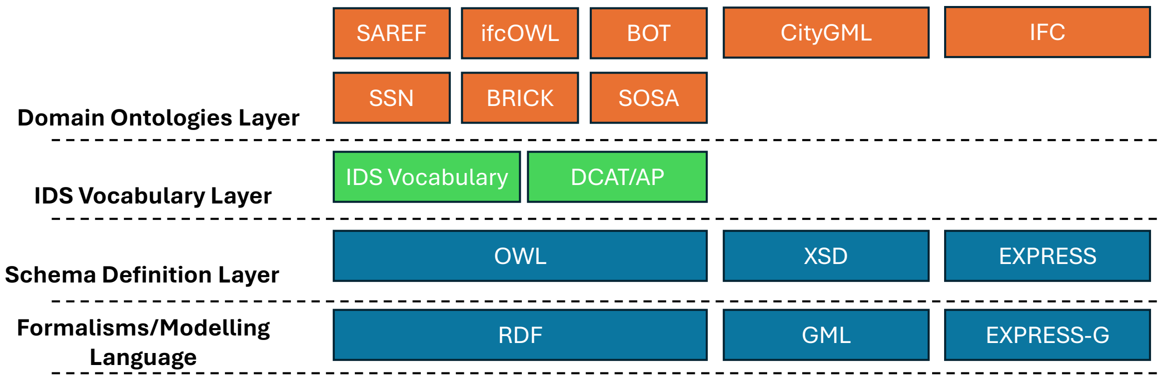

Figure 3: Hierarchy of Ontology setup in an IDS Vocabulary Hub

```
        ids:instance sbm:artifact131415 .

    # Contract offer specifying usage policies
    sbm:contract101112 a ids:ContractOffer ;
        ids:permission sbm:permission161718 .

    sbm:permission161718 a ids:Permission ;
        ids:description "Research_purposes_only"
            ;
        ids:action ids:USE .

    # File reference (points to actual data)
    sbm:artifact131415 a ids:Artifact ;
        ids:fileName "room205_temp_data.json" .

  <!-- Semantic Metadata about the sensor -->
    # Describes the sensor
  sbm:tempSensor001 a sosa:Sensor ;
        sosa:observes sosa:Temperature ;
        sosa:hasLocation sbm:room205 .

    # Spatial context from IFC ontology
    sbm:room205 a ifc:IfcSpace ;
        ifc:longName "Room_205" .
```

A data consumer can preview available data and then choose whether to join the data space or build a connector to receive it. The metadata broker does not store full vocabularies, but references to those available in the vocabulary hub. To keep data descriptions accurate, the vocabulary hub checks them using rules from ontologies and SHACL. For example, a sensor self-description using SOSA and IFC can be validated to ensure that `sosa:hasFeatureOfInterest` links to a valid IFC element. This helps detect incomplete or corrupted metadata.

Figure 4 illustrates two different scenarios of using the vocabulary hub to publish self-descriptions within a data space. In scenario A, the case of single-modality data, the provider offers two different datasets, but both are described using a single ontology (e.g. BOT). In this case, the consumer can request data by developing connector logic that aligns with only the BOT ontology (Connector A). In Scenario B, the provider uses multiple ontologies (e.g. ifcOWL and BOT) to describe the same data asset. In such cases, the data provider may or may not describe the data asset based on the ontology for which the data will not be published. This ability to describe a single data asset using multiple ontologies (Scenario B) reflects the multi-modality of data in a data space, allowing different semantic perspectives to coexist and serve diverse stakeholder needs. For example, if the two ontologies are modular and intended for architects and engineers, respectively, the data provider can configure connectors so that the interfaces return different data sets based on the context of the query received. Additionally, issuing a data policy to filter who can access a specific dataset further enhances the sovereignty of the data provider.

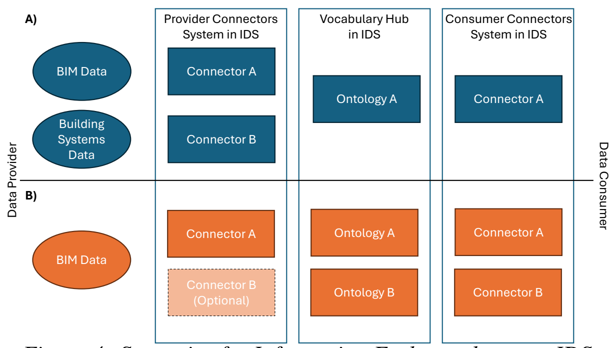

Figure 4: Scenarios for Information Exchange between IDS Connectors and Vocabulary Hub in a data space. A) Two data assets described using one ontology. B) One data asset described using one or more than one ontology.

Discussion on Vocabulary Hubs as a Better Solution to Semantic Data Exchange

Decentralised data exchange and integration are undoubtedly the need of the hour due to increasing concerns about data security, scalability, flexibility, and complexity of integration. Establishing a reliable network of data resources that can scale, as exemplified in the IDS architecture, requires the continued development of robust information systems (Otto et al., 2022). In an economy where all participants contribute and gain value through data sharing, organisations at different levels must embrace data-driven decision-making to improve efficiency and innovation. In this context, this section examines the key challenges of seamless data exchange and evaluates whether the vocabulary hub successfully addresses them. This analysis is particularly relevant for large-scale projects, such as those on the urban scale, where data exchange frameworks are often implemented in a distributed, yet federated manner.

Data Heterogeneity, Multiple Semantic Models and Representations

In modern data ecosystems, data heterogeneity presents a significant challenge to seamless information exchange. Organisations and systems often use multiple semantic models (Steinbuss et al., 2024) to describe the same domain, leading to inconsistencies in how data are structured and understood. For example, in the built environment, ontologies such as ifcOWL, BOT, and SAREF describe buildings differently, making direct interoperability difficult. Beyond semantics, multiple representations (see Figure 3) of the same data, such as RDF graphs, relational databases, and JSON documents, further complicate integration efforts (Pauwels et al., 2017).

A vocabulary hub provides ontology alignment, allowing mappings between different domain-specific models so that participants in a decentralised data space can understand each other's data structures. By exposing standardised APIs, SPARQL endpoints, or data transformation services that convert data into RDF-based formats for semantic querying, the hub ensures semantic consistency when integrating heterogeneous data sources. In the service layer in the architecture, we develop interfaces that use techniques such as label matching to generate equivalence mappings (e.g. `owl: equivalentClass`). The resulting alignments are stored and served through APIs, facilitating semantic interoperability between participants in the data space.

Ontology Reuse and Standardization

Ontology reuse plays a critical role in enhancing data exchange, data consistency, and interoperability. However, many ontologies are not shared in the public domain (Farghaly et al., 2023) or produced with text-based descriptions and therefore lack widespread industry adoption. The IDS rulebook further supports this by emphasising the importance of reusing existing ontologies to prevent fragmentation and redundancy (IDSA-Rulebook, 2025). By adopting widely accepted standards and ontologies (e.g., DCAT, SAREF, BOT), vocabulary hubs support the reuse of existing frameworks, preventing duplication, and ensuring consistency across datasets. Standardised ontology usage can streamline data integration processes and make it easier for organisations to adopt compatible vocabularies and ensure that data exchanged across systems adheres to shared semantics.

Semantic Interoperability

The next major challenge in data exchange is achieving semantic interoperability. Data may be syntactically compatible (i.e., it may follow the same format or schema), but still lack semantic meaning. For example, two systems might exchange data about a 'building', but the definition of a 'building' might differ between systems. This semantic gap creates issues in data integration, analysis, and decision making. In large-scale projects, these interoperability issues become more complex due to the involvement of multiple stakeholders, diverse data models, and cross-domain interactions.

A vocabulary hub offers a potential solution by acting as an intermediary data exchange element or semantic bridge that supports ontology alignment and data translation (Čustović et al., 2023). In addition, dynamic loading of terms in a vocabulary is possible within the IDS framework, thereby solving the interoperability issues of loading a large ontology (Tchouanguem Djuedja et al., 2019; Farghaly et al., 2023). However, its effectiveness depends on the extent to which it facilitates cross-domain mappings.

A vocabulary hub can use mappings to link domain-specific ontologies (e.g. IFC, BOT, SAREF) with standardised vocabularies like IDS, ensuring that each data provider's information is correctly understood by the receiving system. The IDS Information Model supports ontologies based on RDF. In AECO, most vocabularies—except for a few, such as CityGML—are RDF- or OWL-based, which is the industry standard. Such non-RDF based ontologies are, however, required to test the capability of data exchange framework at scale. Hence, a thorough analysis is essential to standardise these ontologies for a data space and to make them available as OWL ontologies.

Support for Cloud-BIM Data Exchange

The study by Afsari et al. (2017) points out limitations in the interoperability of BIM data for cloud-based collaboration. Current interoperability frameworks often rely on syntactic transformations rather than semantic enrichment, which restricts their ability to facilitate seamless data exchange. The vocabulary hub can standardise the cloud-based BIM APIs, conduct automated compliance checks, especially in smart building projects.

Conclusions and Future Work

Data-driven building systems, which are currently operated in silos, require an efficient and standardised method of data exchange. Although centralised approaches to data handling offer a holistic view of information, they also introduce challenges such as scalability limitations and inflexibility in supporting diverse data formats. Decentralised data exchange frameworks, such as IDS, aim to address these challenges by enabling secure, interoperable, and flexible data sharing mechanisms.

Ontologies play a crucial role in achieving semantic data exchange and interoperability, particularly in the AECO domain. As business requirements evolve, semantic models must also adapt to the increasing needs for data complexity and integration. Within the IDS framework, the vocabulary hub emerges as a key enabler, offering potential solutions to challenges related to scalability, integration complexity, security, and cross-organisational interoperability.

Several important research questions remain open or are only partially addressed in this paper: management and governance of vocabularies (e.g., storing multiple versions, defining ownership for updates, validation mechanisms, and self-description), and how to implement rules for ontology mapping and alignment. The possibility of dynamic namespace updates and role-based access control over vocabularies is yet to be explored.

Two additional challenges require further investigation: handling large file sizes and overcoming unidirectional data exchange limitations. While the current IDS architecture supports dynamic retrieval of vocabulary terms based on specific data requests, IDS-RAM lacks optimised mechanisms for large-file streaming. This highlights the need for segmentation techniques to transfer large data sets in smaller packets. Hybrid querying strategies should also be explored to enable access and reasoning over vocabularies defined in XSD, JSON Schema, or EXPRESS.

In a collaborative data space, a data consumer can also act as a provider, requiring real-time data synchronisation. However, it remains unclear how this collaboration, such as in cloud-based BIM platforms, fits into IDS-based frameworks. Future research should focus on handling large file sizes, overcoming unidirectional data exchange limitations, and enabling state-dependent real-time data synchronisation across decentralised building data ecosystems.

Acknowledgments

This work is funded by the European Union under the Horizon Europe research and innovation programme - Project WILSON [Grant Agreement ID: 101147267]. The authors thank the project consortium partners for conducting knowledge-sharing workshops on data spaces.

References

Afsari, K., Eastman, C., and Shelden, D. (2017). Building Information Modeling Data Interoperability for Cloud-Based Collaboration: Limitations and Opportunities. International Journal of Architectural Computing, 15(3):187–202.

Chamari, L., Petrova, E., and Pauwels, P. (2023). An end-to-end implementation of a service-oriented architecture for data-driven smart buildings. IEEE Access, 11:61–81.

Costin, A., Ouellette, J. W., and Beetz, J. (2022). Building Product Models, Terminologies, and Object Type Libraries. In Buildings and Semantics, pages 3–24. CRC Press.

Čustović, I., Soman, R. K., Pauwels, P., and Hall, D. M. (2023). Towards Integrated Digital Twins for Construction and Manufacturing. In Proceedings of the 2023 European Conference on Computing in Construction and 40th International CIB W78 Information Technology for Construction Conference, volume 4, pages 0–0. European Council on Computing in Construction (EC3).

David, R., Ivanov, P., and Alexiev, V. (2024). Raising the Role of Vocabulary Hubs for Semantic Data Interoperability in Dataspaces.

Farghaly, K., Soman, R. K., and Zhou, S. A. (2023). The Evolution of Ontology in AEC: A Two-Decade Synthesis, Application Domains, and Future Directions. Journal of Industrial Information Integration, 36:100519.

IDSA (2025). International Data Spaces – Reference Architecture Model. `https://docs.internationaldataspaces.org/ids-knowledgebase/ids-ram-4/layers-of-the-reference-architecture-model/3-layers-of-the-reference-architecture-model`. Accessed January 30, 2025.

IDSA-Rulebook (2025). International Data Spaces. `https://internationaldataspaces.org/idsa-rulebook/`. Accessed January 30, 2025.

Nagel, L., Hierro, J. J., Perea, E., Lycklama, D., Mertens, C., Taillandier, A.-S., Marques, M., Gelhaar, J., Marguglio, A., Ahle, U., et al. (2021). Design Principles for Data Spaces: Position Paper. Technical report, E. ON Energy Research Center.

Otto, B., ten Hompel, M., and Wrobel, S. (2022). Designing Data Spaces: The Ecosystem Approach to Competitive Advantage. Springer Nature.

Pauwels, P., Zhang, S., and Lee, Y.-C. (2017). Semantic Web Technologies in AEC Industry: A Literature Overview. Automation in Construction, 73:145–165.

Richards, M. (2022). Software Architecture Patterns. O'Reilly Media, Inc., Sebastopol, CA, second edition edition.

Steinbuss, S., Turkmayali, A., Van den Berg, W., Berwanger, S. G., Bilbao Arechabala, S., Fernandez, I., Gil Inchaurza, G., Klinker, P., Lähteenoja, V., Müller, A., and Stornebrink, M. (2024). Semantic Interoperability in Data Spaces.

Taboada-Orozco, A., Yetongnon, K., and Nicolle, C. (2024). Smart Buildings: A Comprehensive Systematic Literature Review on Data-Driven Building Management Systems. Sensors, 24(13):4405.

Tang, S., Shelden, D. R., Eastman, C. M., Pishdad-Bozorgi, P., and Gao, X. (2019). A Review of Building Information Modeling (BIM) and the Internet of Things (IoT) Devices Integration: Present Status and Future Trends. Automation in Construction, 101:127–139.

Tchouanguem Djuedja, J. F., Karray, M. H., Foguem, B. K., Magniont, C., and Abanda, F. H. (2019). Interoperability Challenges in Building Information Modelling (BIM). In Enterprise Interoperability VIII: Smart Services and Business Impact of Enterprise Interoperability, pages 275–282. Springer.