

Semantic Discovery and Selection of Data Connectors in International Data Spaces

Danniar Reza Firdausy¹, Patrício de Alencar Silva^{1,2}, Marten van Sinderen¹ and Maria Eugenia Iacob¹

¹ University of Twente, Drienerlolaan 5, Enschede, 7522 NB, The Netherlands

² Rio Grande do Norte State University (UERN) Federal University of the Semi-Arid Region (UFERSA), Mossoró, RN, Brazil

Abstract

Data sovereignty is the right that individuals and organizations own to control the access to and the disclosure of their private and sensitive data. In Europe, the International Data Spaces Association (IDSA) aims to promote this right by proposing technical and organizational guidelines to help companies build trusted data exchange ecosystems. The IDSA suggests the IDS Connectors as software components necessary to enforce data sovereignty on the technical level. Among many possible functionalities, an IDS Connector could enable (1) data exchange between data owners' and data user's Enterprise systems in a standardized communication protocol; (2) data access policy enforcement; and (3) internal data transformation operations, e.g., integration, mapping, or merging. However, software and service providers may start soon offering IDS Connectors with different configurations through multiple platforms on the Web, making the practical adoption of the IDS architectural guidelines more difficult, especially for small and medium enterprises. We propose developing an IDS Connector Store to discover and select IDS Connectors in IDS ecosystems to cope with this issue. The store will operate as a metadata repository to describe the connectors according to contextual information, e.g., the business domain, pricing model, and data access policies enforced. This paper reports on the current state of this research endeavor by providing a threefold contribution. First, it elaborates on research questions, methods, and goals to address the design problem on hand. Second, it presents an ontology requirements specification document highlighting competency questions related to discovering and selecting IDS Connectors in an IDS ecosystem. Last, it provides the first conceptual draft of an ontology for IDS Connectors described in OntoUML posed for discussion among the conceptual modeling community and to guide meaningful and further specification in Web Ontology Language (OWL).

Keywords

International data spaces, IDS, ontology, semantic web, discoverability, data sovereignty

1. Introduction

IT-based platforms have proven their significance in facilitating data-sharing and interoperability among organizations [1]. A few of their benefits perceived by business entities are improving their planning process, enhancing their capability to fulfill large work orders, and stimulating the creation of new business models [2, 3]. Despite these gains, establishing a data-sharing ecosystem has challenges, e.g., conflicting data formats and standards between companies' software systems and the lack of technical enforcement in disclosing sensitive data [4, 5].

Proceedings of the Workshop of I-ESA'22, March 23–24, 2022, Valencia, Spain

EMAIL: d.r.firdausy@utwente.nl (D.R. Firdausy); p.dealencarsilva@utwente.nl (P. de Alencar Silva); m.j.vansinderen@utwente.nl (M. van Sinderen); m.e.iacob@utwente.nl (M.E. Iacob)

ORCID: 0000-0001-9743-9754 (D.R. Firdausy); 0000-0001-6827-1024 (P. de Alencar Silva); 0000-0001-7118-1353 (M. van Sinderen); 0000-0002-4004-0117 (M.E. Iacob)



© 2022 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)

Subsequently, these hurdles in carrying out data exchange led to the advancement of the International Data Spaces (IDS). This initiative is a decentralized and usage policy-enforced data-sharing ecosystem that puts forward trust, security, interoperability, and data sovereignty in mind [6, 7]. IDS grants the participants access to join and share data in the ecosystem through the IDS Connectors. These software components will be critical to enabling secure data exchange between a data provider and data consumer by enforcing usage policies for the data consumer to use, process, and proliferate the shared data [8]. As a gateway between a company's enterprise systems and their partners', the IDS Connectors may soon be offered and supplied by software and service providers in numerous configurations to satisfy diverse sets of participants' needs and capabilities. As a result, the information to request these IDS Connectors will be increased in a scattered manner throughout the Web, hampering the potential participants' ability to discover them and ultimately limiting their adoption of the IDS vision. To cope with this issue, we propose the idea of an IDS Connector Store – a repository of metadata describing the functionality and contextual information about data connectors.

In the meantime, Semantic Web technology has gained more prominence in information sharing. It has been implemented in an increasing variety of contexts in recent years to enhance the discoverability and accessibility of resources on the Web [9]. One of the building blocks that constitute the Semantic Web is Ontology, which is a formal and explicit specification of a concept used to add a layer of metadata to the described resources to define their meaning [10]. This procedure makes the Web more accessible and understandable for more refined search results by software agents in providing information to human agents. Even though considerable research has adopted the Semantic Web approach, the ones that attempt to cope with the widespread proliferation of the IDS Connectors are still minimal. Therefore, we propose the application of Ontology and Semantic Web technology to the development of the IDS Connector Store to facilitate the discoverability and selection process of the IDS Connectors.

The organization of the rest of this paper follows. **Section 2** will address the question that probes the methodological steps required to develop the ontology for the IDS Connectors. Then, **Section 3** elaborates on producing the ontology requirements specification document, highlighting the competency questions related to discovering and selecting the IDS Connectors. Finally, in **Section 4** and **Section 5**, we will present the first conceptual draft of the mentioned ontology described in OntoUML and pose a discussion among the conceptual modeling community to specify the ontology further in Web Ontology Language (OWL).

2. Research methods

The first prominent step in developing an ontology-based software application is the formulation of the ontology requirements specification document (ORSD). In this research, we adopt the scenario-based NeON Methodology, emphasizing the reuse of existing ontological and non-ontological resources in developing the ontology [11]. In addition to the requirements specification activity guidelines, this methodology also provides a template to formulate the ORSD as a filling card that describes the purpose, scope, implementation language, intended end-user, intended uses, requirements, and pre-glossary terms of the ontology under design [12]. The process followed to produce the ORSD based on this methodology will be discussed in more detail in the next section.

This paper aims to produce a preliminary ontology for the IDS Connectors to facilitate their discoverability and selection for the IDS participants. To maintain interoperability with the domain reference ontology, NeON suggests a quick search of knowledge resources for possible reuse during the development. For this purpose, the IDSA has published the IDS Information Model that describes the fundamental concepts of the IDS, covering entities from the participants to the infrastructure components [13]. This IDS Information Model grounds the ontology proposed in this work. The resulting conceptual model is depicted in OntoUML [14]. This model serves as the basis for further implementation into OWL to describe the IDS Connectors and distinguish them with subject-predicate-object triples according to the Resource Description Framework (RDF) format [15]. Through this semantic annotation, several sentences can be formed to explain the IDS Connectors. For instance, company A maintains an IDS Connector X, IDS Connector Y is offered in a flat-rate

pricing model, or IDS Connector Y complies with GS1 standards [16, 17]. As a result, software agents will be able to discover the IDS Connectors that are appropriate to their data exchange demands.

3. IDS connector ontology requirement specification

The requirements specification identifies the ontology's purpose, scope, and implementation language. As presented in Table 1, three main end-users that will take advantage of the knowledge given by the IDS Connector ontology are listed. The business representatives are the first target users due to their interest in spotting potential business opportunities in the current business landscape. For the potential IDS participants, the presence of their partners and the prospect of securing a strategic partnership with other existing participants signal the value accessible to them by participating in the data space. Such a scenario might influence their willingness to consolidate into the IDS ecosystem.

Conversely, the interests of the existing participants can take many forms. One example is to find other prospective partners to engage in strategic information exchange to leverage their value chain performance. The IT representatives will need to further translate these business strategies to the IT implementation strategies by investigating the suitable IDS Connector matching their needs and capabilities. Such a demand leads to concerns about which IDS Connectors fit their business domain or industrial standards adopted for data exchange. In response, software and service providers will be interested in making their IDS Connectors discoverable by external software applications.

Table 1

IDS Connector Ontology Requirement Specification Document

Purpose

To serve the IDS Connector Store as a knowledge base in describing the IDS Connector to guide existing and potential participants of the IDS ecosystem to the relevant IDS Connector.

Scope

The ontology focuses on describing and discovering the IDS Connectors according to contextual information (e.g., the business domain, pricing model, and enforced data access policy) with the granularity represented by the competency questions.

Implementation Language

The ontology is represented in OntoUML, with further translation into OWL.

Intended End-Users

- User 1. Business representatives of potential and existing IDS participants
- User 2. IT representatives of potential and current IDS participants
- User 3. Software and Service Providers who develop and supply IDS Connectors
- User 4. Scholars who are keen to explore the ontology's knowledge representation capabilities

Intended Uses

- Use 1. Software and Service Providers publish their offered IDS Connectors' metadata on the IDS Connector Store to make their offered IDS Connectors discoverable.
- Use 2. Business representatives search for IDS-compliant partners operating in the same business domain, complying with common standards, etc.
- Use 3. IT representatives search for IDS Connectors that match their needs and capabilities.
- Use 4. Scholars search for, use, and import the ontology into their proof-of-concept IDS implementations.

Ontology Requirements

Non-Functional Requirements

- NFR 1. The ontology must at least use English
- NFR 2. The ontology must comply, reuse and integrate with the existing IDS Ontology specified under the IDS Information Model.

Functional Requirements: Competency Questions

CQG1. IT Representatives

- CQ 1. What software provider offers IDS Connectors?
- CQ 2. Which IDS Connectors are developed for a specific business domain?
- CQ 3. Which IDS Connectors are complying with a particular standard?
- CQ 4. Which IDS Connectors are offered in this pricing model?
- CQ 5. Which IDS Connectors support these data usage agreements?
- CQ 6. Which IDS Connectors are developed using this application development framework?
- CQ 7. Which IDS Connectors are offered in this deployment context?

CQG2. Business Representatives

- CQ 8. Which IDS participants use a particular IDS Connector from a specific Software Provider?
- CQ 9. Which IDS participants operate in a particular business domain?
- CQ 10. Which IDS participants comply with a particular standard?

Pre-Glossary of Terms

Terms from Competency Questions & Frequency

- | | | |
|---------------------|-------------------|------------------------|
| - IDS Connector | - Business Domain | - Data Usage Agreement |
| - Participant | - Standards | - Technology |
| - Software Provider | - Pricing Model | - Deployment |

Objects and Terms for Answers

- | | | |
|--|---|--|
| - Gatewise IDS Connector, Supplydrive IDS Connector, TradeCloud IDS Connector. | - Transport Logistics, Glass Manufacturing, Steel Manufacturing, etc. | - Delete After Interval Agreement, Connector-restricted Agreement, Logging Agreement, etc. |
| - Vandaglas B.V., Van Egmond Groep, Meijer Metal | - OTM, GS1, EDI4STEEL, etc. | - Java, Spring Boot, JavaScript, NodeJS, VueJS, Python, etc. |
| - ECI Software Solutions, Tradecloud | - Flat Rate, Freemium, Pay per User, Pay per Feature, etc. | - On-Premise, Cloud: SaaS, etc. |

The functional and non-functional requirements specification then follows the Ontology Engineering activity. The ontology development work on the IDS itself was initiated by the IDSA by publishing the IDS Information Model. This work grounded the development of this IDS Connector ontology, putting it as a prominent non-functional requirement. On top of this, several functional requirements in the form of competency questions lead to the semantic-enabled discovery and selection process of the IDS Connectors. These questions are grouped by considering the immediate interests of the relevant end-users. Frequent terms are extracted from the competency questions, leading to the enumeration of objects for answering the end user's query. We instantiate the entities listed in Table 1 above by referring to the literature [5, 18], industrial standards [16, 17, 19], IDSA documentation and publications [8, 13, 20], and the publication of the Smart Connected Supplier Network (SCSN), one of the IDS forerunners in the Dutch manufacturing supply chain [21, 22].

4. Preliminary IDS connector ontology conceptual model

Using the IDS Reference Architecture Model (RAM) and the IDS Information Model (IM) as a starting point, we have identified several concepts relevant to answering the CQs above, namely the Participant and the Connector concepts [8, 13]. As shown in Figure 1, we identify the former concept as the IDS Actor and extend it further into two specializations. The Core IDS Actors refers to the participants who either own and provide or request and use data. Meanwhile, the IDS Supporting Actors are associated with parties that ensure the continuation of the data-sharing ecosystem. The Software and Service Provider carries this duty by providing essential components to participate in the data space.

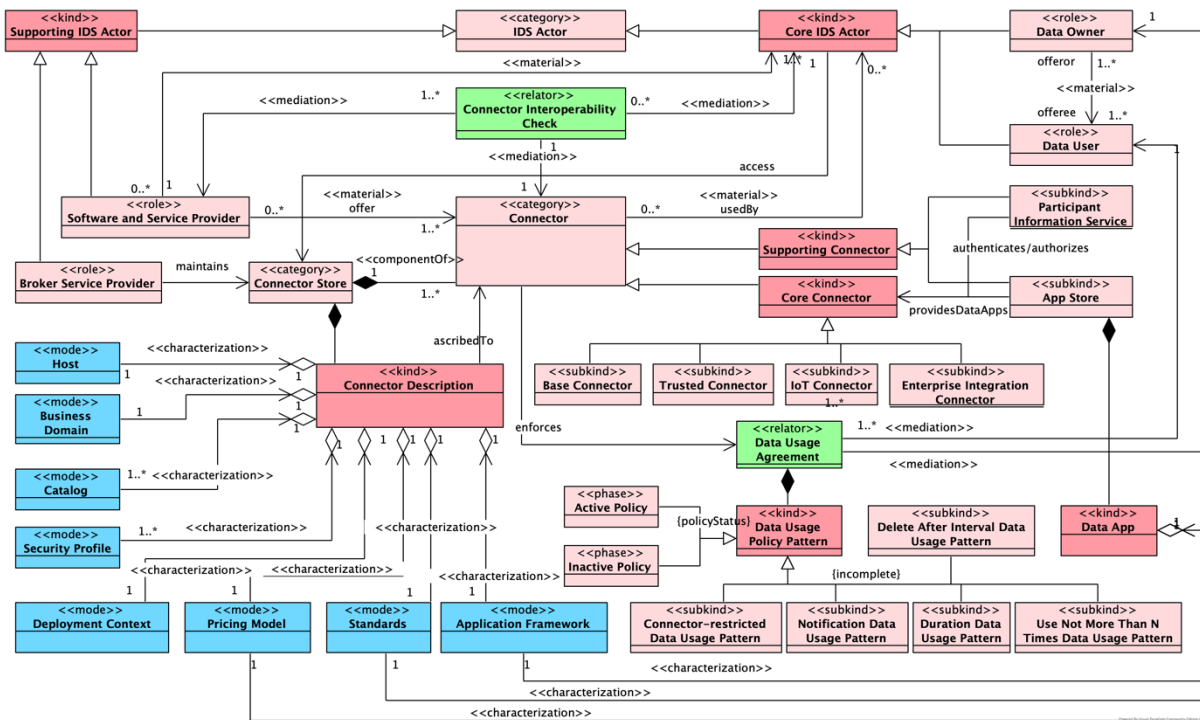


Figure 1: Preliminary IDS Connector Ontology Conceptual Model

Meanwhile, the Broker Service Provider supports the core actors with the function to look up for the other actors and the Connector used by the different actors through the functionality offered by the IDS Connector Store. In addition, the Supporting IDS Actor also covers other roles, such as the Clearing House and Identity Provider. However, as the IDS RAM describes, these roles can be assumed by the same organization that takes the part of the Broker Service Provider [8].

The IDSA expresses the Connectors from several different perspectives. On the one hand, the IDS IM describes the concept of a Connector to be the generalization of the Base Connector, Trusted Connector, App Store, and Participant Information Service [13]. Here, we distinguish these Connectors into the Core Connector and Supporting Connector, each used by the corresponding type of role. The IDS RAM justifies this distinction by describing that the functions falling into the supporting category, i.e., the App Store Provider, Broker Service Provider, and Identity Provider, rely on the Connector technology to carry out their functions [8]. On the other hand, the IDS RAM also characterizes the Connector from its Deployment Context, Security Profile, Catalog, and Host. The Deployment Context is designated as the Connector's deployment environment, i.e., on-premises or cloud-based. Security Profile explicates the Connector's capability to enact a secure data exchange and processing environment. Host signals the communication protocol that the Connector supports to expose resources, i.e. HTTPS URLs, MQTT topics, etc. Whereas, the Catalog facilitates the participant discovery in the ecosystem based on the digital resources that the Connector provides or consumes.

We extend the Connector concept with additional properties to facilitate its discovery and selection. The Business Domain describes the context where the Connector is developed to be specialized. Standards refer to the criteria to which the connector complies. The Pricing Model implies how the end-users are expected to pay for the Connector's usage and acquisition. The Application Framework informs which technology stacks are used to develop and support the Connector's runtime. Finally, the Data Usage Agreement is understood as the contract composed of the Data Usage Policy Pattern and agreed by the interacting Core IDS Actor to govern the data usage. As of now, five types of data usage patterns are supported by the Connector, and more designs may be added in the future.

5. Conclusion

This paper introduced a preliminary ontology conceptual model to describe IDS Connectors that will serve as a façade of a Connector Store. Since this ontology is still under development, we plan to refine it further to accommodate more relevant competency questions to facilitate the discovery of IDS Connectors and IDS actors. The next step is to translate this model into OWL representation for semi-automated machine reasoning. Then, we populate the resulting representation with instances of IDS Connectors and IDS Actors and load it into a triple store such as TriplyDB to make it publicly available [23]. Then, we plan to evaluate the published ontology by translating the competency questions into SPARQL queries to verify the ontology's correctness, consistency, and completeness. Hence, we plan to validate the utility of the ontology by assessing the relevance of the posed competency questions with expert opinion and the end-users identified in Table 1, besides verifying if the answers returned from the SPARQL correspond to end-users expectations. Lastly, the completed and published ontology will be operationalized into a proof-of-concept implementation of the IDS Connector Store by connecting the triple store with a user-interfacing application to support the discovery and selection of IDS Connectors in an IDS-compliant data-sharing ecosystem.

6. Acknowledgements

This research is financially supported by the Dutch Ministry of Economic Affairs and co-financed via TKI DINALOG and NWO. The CLICKS project has granted funding for this work (grant no. 439.19.633). CLICKS is the acronym for Connecting Logistics Interfaces, Converters, Knowledge, and Standards. The authors thank the involved consortium partners for their support and the anonymous reviewers for their constructive feedback.

7. References

- [1] M. L. Markus, Q. N. Bui, Going concerns: The governance of interorganizational coordination hubs, *Journal of Management Information Systems* 28 (2012) 163-198. doi: 10.2753/MIS0742-1222280407
- [2] M. Banek, D. Juric, D. Pintar, Z. Skocir, M. Vranic, B. Vrdoljak, E-business infrastructure for supporting the integration of tourist services, in: 2008 50th International Symposium ELMAR, IEEE, New York, 2008, pp. 289-292.
- [3] X. Wang, C. Zhang, Y. Jin, X. Zhao, CPSP: A Cloud-based Production Service Platform Supporting Co-Manufacturing of Cross-Enterprise, in: 2018 IEEE 22nd International Conference on Computer Supported Cooperative Work in Design (CSCWD), IEEE, New York, 2018, pp. 455-460, doi: 10.1109/CSCWD.2018.8465354.
- [4] A. Braud, G. Fromentoux, B. Radier, O. Le Grand, The road to European digital sovereignty with Gaia-X and IDSA, *IEEE Network* 35 (2021) 4-5. doi: 0.1109/MNET.2021.9387709.
- [5] I. Lopes-Martínez, L. Paradela-Fournier, J. Rodríguez-Acosta, J. L. Castillo-Feu, M. I. Gómez-Acosta, A. Cruz-Ruiz, The use of GS1 standards to improve the drugs traceability system in a 3PL Logistic Service Provider, *DYNA* 85 (2018) 39-48.
- [6] S. Dalmolen, H. Bastiaansen, E. Somers, S. Djafari, M. Kollenstart, M. Punter, Maintaining control over sensitive data in the Physical Internet: Towards an open, service oriented, network-model for infrastructural data sovereignty, 2019. URL: <https://repository.tno.nl/islandora/object/uuid%3Ab2e6952a-06ed-46e1-b186-fc25932b28c3>
- [7] B. Otto, M. Jarke, Designing a multi-sided data platform: findings from the International Data Spaces case, *Electronic Markets* 29 (2019) 561-580. doi: 10.1007/s12525-019-00362-x.
- [8] International Data Spaces Association, IDSA Reference Architecture Model Version 3.0, 2019. URL: <https://internationaldataspaces.org/wp-content/uploads/IDS-Reference-Architecture-Model-3.0-2019.pdf>
- [9] K. Janowicz, F. Van Harmelen, J. A. Hendler, P. Hitzler, Why the data train needs semantic rails, *AI Magazine* 36 (2015) 5-14.
- [10] S. Salma, M. Bouneffa, C. Habiba, Ontology and Semantic Web in Logistic Applications: State of the Art, in: 2019 7th Mediterranean Congress of Telecommunications (CMT), IEEE, New York, 2019, pp. 1-4. doi: 10.1109/CMT.2019.8931374.

- [11] A. Gómez-Pérez, M. C. Suárez-Figueroa, NeOn methodology for building ontology networks: a scenario-based methodology, 2009. URL: https://oa.upm.es/5475/1/INVE_MEM_2009_64399.pdf
- [12] M. C. Suárez-Figueroa, A. Gómez-Pérez, B. Villazón-Terrazas, How to write and use the ontology requirements specification document, in: OTM Confederated International Conferences" On the Move to Meaningful Internet Systems, Springer, Berlin, 2009, pp. 966-982.
- [13] IDSA. The International Data Spaces (IDS) Information Model, 2021. URL: <https://github.com/International-Data-Spaces-Association/InformationModel>.
- [14] G. Guizzardi, Ontological foundations for structural conceptual models, 2005. URL: https://ris.utwente.nl/ws/portalfiles/portal/6042428/thesis_Guizzardi.pdf
- [15] T. Berners-Lee, J. Hendler, O. Lassila, The semantic web, Scientific american 284 (2001) 34-43.
- [16] GS1, GS1 Transport & Logistics, 2021. URL: <https://www.gs1.org/industries/transport-and-logistics>.
- [17] OpenTripModel. What is the Open Trip Model?, 2021. URL: <https://www.opentripmodel.org/page/about>.
- [18] W. Bol Raap, M.-E. Iacob, M. v. Sinderen, S. Piest, An architecture and common data model for open data-based cargo-tracking in synchromodal logistics, in: OTM Confederated International Conferences" On the Move to Meaningful Internet Systems, Springer, Berlin, 2016, pp. 327-343.
- [19] INAD Industrie Software B.V., EDI4STEEL, 2022. URL: <https://www.edi4steel.eu/about/>.
- [20] IDSA. Dataspace Connector, 2021. URL: <https://github.com/International-Data-Spaces-Association/DataspaceConnector>.
- [21] C. Stolwijk, F. Berkers, Scalability and agility of the Smart Connected Supplier Network approach, 2020. URL: <https://repository.tudelft.nl/islandora/object/uuid%3A36745cb0-3d5f-4f79-9034-93e02e80529c>
- [22] SCSN, Smart-Connected Supplier Network (SCSN) Addressbook, 2020. URL: <https://broker.ids.smart-connected.nl/#home>.
- [23] TriplyDB, The Netowrk Effect for your Data, 2022. URL: <https://triplly.cc/>.