Establishing Semantic Interoperability across data spaces: a solution for sharing vocabularies.

TNO Publiek › TNO 2024 R10911
17 januari 2024

ICT, Strategy & Policy
www.tno.nl

TNO 2024 R10911 – 17 januari 2024

Establishing Semantic Interoperability across data spaces: a solution for sharing vocabularies.

Auteurs	Jelte Bootsma, Jan-Pieter Wijbenga, Linda Oosterheert, Michiel Stornebrink, and Wouter van de Berg
Rubricering rapport	TNO Publiek
Titel	TNO Intern
Rapporttekst	TNO Intern
Aantal pagina's	4 (excl. voor- en achterblad)
Aantal bijlagen	0
Projectnaam	Centre of Excellence for Data Sharing and Cloud
Projectnummer	060.54832

Alle rechten voorbehouden
Niets uit deze uitgave mag worden verveelvoudigd en/of openbaar gemaakt door middel van druk, fotokopie, microfilm of op welke andere wijze dan ook zonder voorafgaande schriftelijke toestemming van TNO.

© 2024 TNO

Inhoudsopgave

1 Executive Summary

To unlock the full potential of data sharing, data spaces are increasingly gaining traction. Sharing data across different data spaces holds even more value, particularly in the context of achieving EU climate goals. The European Strategy for data summarizes EU's ambition of achieving data spaces that are interoperable within the data space itself and interoperability amongst data spaces – a federation of data space [1]. However, when data needs to be shared across data spaces, it is even more challenging to ensure semantic interoperability, as each data space develops their own vocabularies and tools. A standardized solution is required to bridge the semantic interoperability gap across data spaces.

This paper focuses on a standardized approach for exchanging vocabularies across data spaces, allowing publishers to describe and share vocabularies. Currently, datasets and data services can be published and shared in various catalogues, enhancing their discoverability within federated data spaces. Recognizing that vocabularies can be treated as datasets, this paper proposes to utilize the Data Catalogue Vocabulary Application Profile (DCAT-AP) [2] as a standardized model for describing and exchanging vocabularies between data spaces. This approach aims to increase vocabulary discoverability, facilitate federated searches, and maximize the reuse of vocabularies across different sectors.

DCAT-AP emerges as a standard for describing datasets and data services within a data space. Instead, this paper explores the applicability of DCAT-AP as a candidate standard for providing a common way for data spaces to describe and share their vocabularies. This ensures a uniform interpretation of vocabularies across multiple data spaces, supporting semantic interoperability and unambiguous data sharing between two data spaces. The exploration of this concept includes a proof of concept in an open-source vocabulary hub, called Semantic Treehouse [3].

The uniform approach for sharing vocabularies presented in this paper serves as a starting point for achieving semantic interoperability across data spaces. It allows data space participants to explore vocabularies from other hubs as if they were present in their own, enhancing the discoverability of vocabularies within federated data spaces. This discoverability acts as the foundation for negotiating a common vocabulary or mapping between two different vocabularies.

2 Introduction

In a world where information needs to flow between different systems, organizations, and sectors, data spaces are developed to facilitate the exchange of data. In a data space it is crucial to have a clear understanding of what the data means to prevent mistakes and miscommunication while sharing data. When data is shared with an unambiguous and clear understanding this is called *semantic interoperability* [4]. To achieve this, data space participants need to express their data offering or data need using a common vocabulary. In this paper, we use the term common vocabulary for any specification that establishes a standard language for consistently describing, interpreting and annotating data by offering a structured sets of terms, concepts, and definitions. These vocabularies are like dictionaries that help a data provider and data consumer speak the same language when it comes to exchanging data. For example, consider a JSON schema specification that clarifies the structure and usage of a particular JSON document. The main responsibility for this common language lies with an intermediary role called a *vocabulary provider*. This party manages and offers vocabularies that can be used to annotate and describe data.

2.1 Vocabulary hub

To make vocabularies findable, accessible, and usable for the data space participants, a data space deploys a *vocabulary hub*. According to the IDS RAM 4 [5], the *vocabulary hub* is a service that stores, maintains, and publishes the vocabularies and enables collaborative management of the vocabularies. It is a service supporting vocabulary publication, editing, browsing, and maintenance. The vocabularies itself semantically describe the data that is exchanged between data space participants. Beyond accessibility, it fosters collaboration on vocabularies through features such as version management and a co-creation process. In essence, a vocabulary hub acts as a hub for efficiently managing and publishing the vocabularies used to specify the meaning of data within and across data spaces.

2.2 Federated data spaces

Data spaces are emerging to enable collaboration, by the exchange and sharing of data. Data spaces are currently being developed in numerous sectors and regions, and by individual consortia. To unleash the true potential of data sharing, there is a growing need to enable the exchange of data across different data spaces. Being interoperable on all four key layers of interoperability as defined by the European Interoperability Framework (EIF) [6] is called the *federation of data spaces*. The development of data spaces and the federation of data spaces is gaining a lot of attention and is prominently part of the plans of the European Commission. The European Strategy for Data strives for smooth cooperation amongst different data spaces and effective sharing of information. For example, this can be seen in sharing carbon footprint data to support achieving European climate goals. To make this work, it is essential to ensure that data spaces in the variety of sectors that have carbon footprint data available can, mutually exchange data. To enable seamless data exchange, the data and thus the data spaces should be interoperable on the legal, organizational, semantic, and technical level [6].

2.3 Semantic interoperability in federated data spaces

One of the significant challenges for federated data spaces is ensuring semantic interoperability, as different data spaces develop their unique semantics and structures. The challenge at hand lies in the necessity for a standard that establishes mutual understanding between different data spaces. As vocabularies are stored in vocabulary hubs within a data space, multiple data spaces implies multiple vocabulary hubs and a multitude of vocabularies. The existence of different vocabulary hub implementations results in vocabularies not being accessible across different data spaces, and thus not being used in other data spaces. Current vocabulary hub implementations lack the capability to share their content (the vocabularies they store and publish) in a standardised way with components within or across data spaces, such as other vocabulary hubs. This limitation arises from the absence of a standard to disclose vocabulary hub content between data spaces.

2.4 Use cases

Standardizing the exchange of vocabularies holds significant relevance for several use cases that aim to create interoperability among data spaces. Our vision and proof of concept find relevance in at least the following use cases:

- The federation between operational data spaces in respectively the German automotive (Catena-X [7]) and Dutch high-tech supply chains (SCSN [8]). Such a federation will lead to the emergence of new businesses and an environment where both data space can learn from each other.
- The federation between two vocabulary hubs Smart Data Models [9] and TNO's implementation of a vocabulary hub, known as Semantic Treehouse [3] Facilitating the exchange of vocabularies among vocabulary hubs will enhance the discoverability of vocabularies, thereby encouraging the reuse of vocabularies within respective domains.
- To work with flex workers in the construction sector, it is important for the construction sector and the flexible staffing industry to become semantically interoperable. This is a collaboration between two Dutch operational data spaces: Ketenstandaard [10] in the construction sector and SETU in the flexible staffing industry [11].

Each use case strives to create interoperability between data spaces, allowing data exchange to reach their common goal. The standardised solution we describe to exchange vocabularies fosters semantic interoperability in these use case.

2.5 DCAT-AP

For realizing our approach and a proof of concept, a standard is needed for describing and exchanging vocabulary hub's content. The chosen standard must be machine-readable, expressed as an open standard, and well-documented, while also enhancing discoverability and visibility of vocabularies at a high level.
In this paper, the Data Catalogue Vocabulary Application Profile (DCAT-AP) [2] is used as a candidate standard, as it is meant to describe datasets and metadata in the context of data portals. DCAT-AP is a specification based on the Data Catalogue vocabulary (DCAT) [12] for describing public sector datasets in Europe. The definitions of datasets and data services in DCAT-AP are broad and inclusive, aiming to embrace data types arising from diverse

communities. For this reason, DCAT-AP also enables providers of vocabularies to describe their vocabularies in a standardized manner. In this paper, we use the term data to refer to datasets in DCAT-AP.
Moreover, this standard meets our criteria for a standardized solution, being an open standard and widely adopted with comprehensive documentation. It also facilitates discoverability by describing generic information about the vocabulary hub's content and enables references to retrieve the contents. Additionally, this standard is widely used in diverse data spaces specifications, simplifying the integration between some data space components.

2.6 Purpose and scope

This paper aims to illustrate that vocabularies can be viewed as data and exchanged in a standardized manner across data spaces. For data spaces seeking to establish federation, this enhances the discoverability of vocabularies, serving as the foundation for negotiating a common vocabulary or mapping between vocabularies of two data spaces.
The current focus is on realizing a standardized manner for the exchange of vocabularies between data spaces, marking the initial step towards semantic interoperability. We propose a vision and a proof of concept for achieving this. While there are other relevant topics crucial for semantic interoperability, such as reaching consensus on vocabulary usage and exchange methods, this paper outlines the starting point for exchanging and reusing vocabularies.

2.7 Structure of the paper

The next chapter describe four types of (meta-)data illustrating the types of data related to a vocabulary hub. Chapter 4 dives into the IT architecture design for a vocabulary hub, and how to annotate vocabularies using DCAT-AP. Chapter 5 illustrates a proof of concept on how to annotate and share vocabularies through DCAT-AP. The specific implementation of a vocabulary hub Semantic Treehouse is used as an environment for this proof of concept. The last chapter describes the conclusions and any future work.

3 Semantic Interoperability in Data Spaces: 4 types of data

Vocabularies are stored in a vocabulary hub and are accessible and usable within a data space. To share vocabularies across vocabulary hubs and in federated data spaces, we consider four types of data.
This section offers our perspective of these four types of data used within data spaces and vocabulary hubs. We use these types to identify where standardization is required for effective vocabulary exchange and how a standard like DCAT-AP can be applied in this context. Below a definition and a simple example for each data type is given. We introduce on what data types DCAT-AP can be applied, but the actual application of DCAT-AP is more extensively described in chapter 4.

Table 1: 4 quadrants for the 4 types of data and their description

Metadata on the data *Descriptive information about the data.*	**Metadata on the vocabulary** *Descriptive information about the vocabulary*
Data *Primary data transmitted, e.g., product data.*	**Vocabulary** *Defines the semantics and structure of data, such as a JSON schema.*

3.1 Data

- **Definition:** The data represents the primary data that is exchanged between participants of data spaces. The actual data is not presented in a vocabulary hub.
- **Example:** Consider the following JSON representation of a product, including a product's name, price, brand, and stock quantity.

```
1. {
2.   "name": "Laptop Ultra-17",
3.   "priceEUR": 899.99,
4.   "brand": "TechCo",
5.   "stockQuantity": 25
6. }
```

Code example 1: JSON representation of product data

3.2 Metadata on the data

- **Definition:** Metadata refers to data that provides information about other data. Metadata about primary data is additional data that provides information about the data itself. Metadata serves to facilitate the management, discovery, and understanding of data. Just as a normal language, besides the words in a dictionary it also has additional information.
- **Example:** Consider the following JSON representation of metadata about the product data. This metadata provides insights into the data presented, such as its title, description, and the date it was established, with attributes like the "title," "description," "author," and "createdDate."

```
{
  "title": "Product Information",
  "description": "A JSON representation including product information.",
  "author": "XYZ Corporation",
  "createdDate": "2022-10-06"
  "version": "1.0",
  "language": "en-US"
}
```

Code example 2: a JSON representation of the data's metadata

3.3 Vocabulary

- **Definition:** A 'vocabulary' is any specification that establishes a language for consistently describing, interpreting and annotating data by offering a structured sets of terms, concepts, and definitions [13, 5].
  The concept of 'vocabulary' is related to but distinct from the concept of 'data model'. The latter refers to an abstraction that describes the structure of data, how it is organized and how different parts relate to each other. For this paper, we consider data models like JSON Schema definitions as a kind of lightweight vocabulary. This is because schemas do define terms (such as field names and types).
  In practice, a vocabulary could also include descriptive information (metadata) about itself. However, we distinguish between the vocabulary content and its metadata in the following subsection.
  Please note that a vocabulary can also be considered as data that can be exchanged. However, it differs from the actual data that is being shared amongst a data provider and a data consumer in the data space (see the data type above). Instead, a vocabulary is not exchange between a data provider and data consumer, but the vocabulary as data is being shared amongst vocabulary hubs.
- **Example:** Consider the following JSON Schema representation of a vocabulary that precisely defines the structure and order of the product data. This vocabulary specifies that "name" and "brand" are strings, "price" is a floating-point number, "stockQuantity" is an integer, and all properties are mandatory to be present. It provides the guidelines for organising the data that is transmitted to ensure the receiver of the data understands what the data is about. Note that this message allows sending of product information for a single product.

```
9   ...
10. {
11. "type": "object",
12. "properties": {
13. "name": {
14. "type": "string"
15. },
16. "priceEUR": {
17. "type": "number"
18. },
19. "brand": {
20. "type": "string"
21. },
22. "stockQuantity": {
23. "type": "integer"
24. }
25. },
26. "required": ["name", "price", "brand", "stockQuantity"]
27. }
```

Code example 3: The part of a JSON schema that defines the structure of data.

3.4 Metadata on the vocabulary

- **Definition:** Metadata on vocabularies represents additional data that provides information about the vocabulary itself. Metadata serves to facilitate the management, discovery, and understanding of data, in this case a vocabulary as data.
  As vocabularies are considered as data, the metadata on the vocabulary, similarly to metadata type in 3.2 can be described and standardized using DCAT-AP.
- **Example:** Consider the following JSON representation of metadata about the vocabulary above. This metadata includes descriptive details such as the vocabularies' "title," "version", "description," "author," and "createdDate". The "$id" identifier identifies this specific product schema and is important when we need to refer to this schema in any context. The "$schema" provides information on how to validate whether this given schema is correct.

```
1.  {
2.    "$schema":"https://json-schema.org/draft-07/schema",
3.    "$id": "https://example.org/product-v1.schema.json",
4.    "title": "ProductSchema",
5.    "version": "1.0",
6.    "description": "Schema for basic product data.",
7.    "author": "John Smith"
8.    "createdDate": "2023-10-06",
9.    …
```

Code example 4: A JSON representation of the schema's metadata

4 Vocabulary Hubs in Federated Data Spaces: an architecture

In the previous chapter we introduced 4 types of data used within data spaces and vocabulary hubs. This chapter addresses how a vocabulary hub can effectively function in a federation of data spaces.
We present the architecture of a vocabulary hub in a federated data space and illustrate where the four data types fit within the architecture. Furthermore, we explore annotating vocabulary metadata in DCAT-AP for the exchange and reuse of vocabularies across various data spaces.

4.1 Components of a vocabulary hub

A vocabulary hub provides users with a clear overview of available vocabularies and how these vocabularies can be effectively utilized. The key components of a vocabulary hub include:

- **Vocabulary creation and editing component;** offering flexibility for starting from scratch or integration of existing vocabularies.
- **Vocabulary repository;** providing storage of all the distributions of vocabularies: any specification that can be used to consistently describe data.
- **Catalogue component;** enabling metadata descriptions and easy access to vocabularies within the vocabulary repository, facilitating their discoverability and reuse.

Typically, the development of vocabularies is organized by business communities and is delegated to standards development organizations (SDOs). A vocabulary hub assists users in creating and publishing these vocabularies. It offers various functionalities for ***vocabulary creation***, whether by reusing existing vocabularies or starting from scratch. All the distributions of vocabularies are stored within the ***vocabulary repository*** of the vocabulary hub.
To ensure the utilization and reusability of these vocabularies, a vocabulary hub provides functionalities for incorporating metadata about each vocabulary. This collection of metadata improves the discoverability of each vocabulary. All the metadata and a link to the vocabularies within the vocabulary repository are published in the ***catalogue component*** of the vocabulary hub.
Figure 1: Components of a vocabulary hub illustrates the relationship between the data types and the repositories of a vocabulary hub. The vocabulary hub is not involved in the actual data transaction between a data provider and a data consumer. Instead, the vocabulary hub only includes the vocabularies to which the data in the transaction must adhere. The catalogue component assists users in easily finding their desired vocabulary. Once the appropriate vocabulary is found by exploring the metadata, users can retrieve the specifications and details on the vocabulary itself from the vocabulary repository. Both the vocabulary repository and the catalogue are usually accessible via an API endpoint.

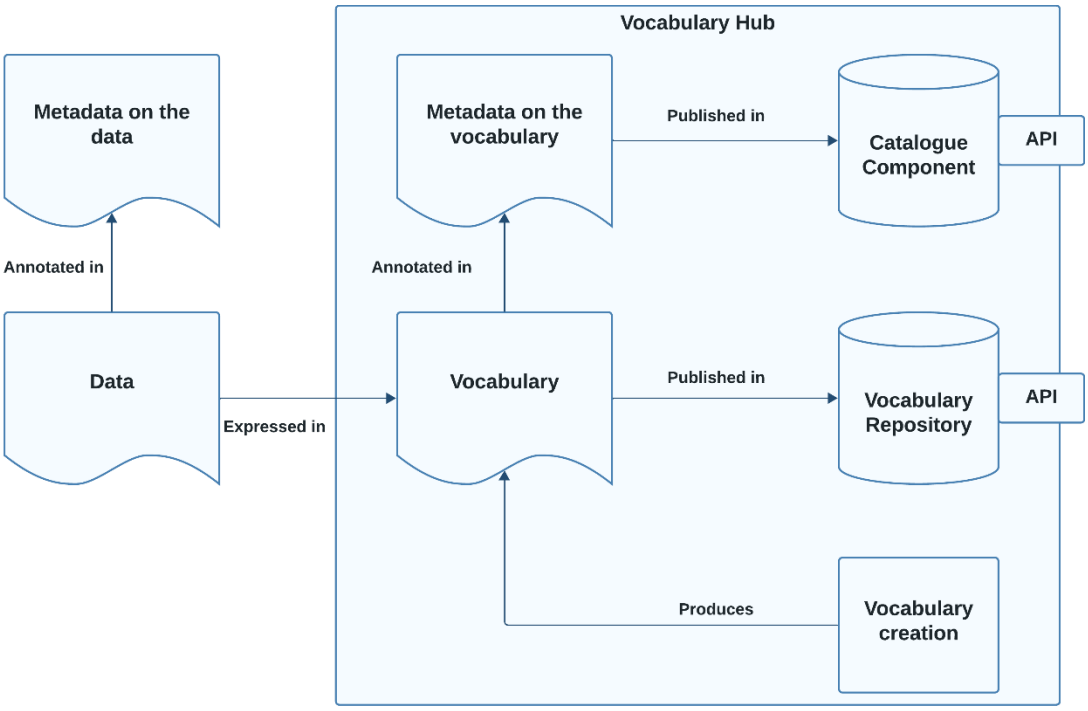

Figure 1: Components of a vocabulary hub

4.2 DCAT-AP for vocabulary hubs

As introduced before, DCAT-AP can be used to facilitate the exchange of vocabularies between different vocabulary hubs by standardizing the metadata of vocabularies. By encouraging all vocabulary hubs to adopt DCAT-AP for metadata descriptions on vocabularies, we unlock the potential for federated searches for vocabularies across diverse vocabulary hubs. This means that when every vocabulary hub consistently employs DCAT-AP for describing metadata, it ensures a standard method for sharing vocabulary hub content within and across data spaces.
However, most vocabulary hubs have their own approach for describing metadata. Some hubs may not even document metadata, while others rely on other existing standards, like the IDSA information model. By adopting DCAT-AP to describe metadata, sharing vocabularies among different hubs in various data spaces becomes a reality.

4.2.1 Describing vocabularies using DCAT-AP

The central notion in DCAT-AP is a Dataset, described as "a collection of data, published or curated by a single source, and available for access or download in one or more formats." (source: DCAT-AP specification 2.1.0). A Data Catalogue on the other hand is described as "a catalogue or repository that hosts the Datasets or Data Services being described". The definition of datasets and catalogues in DCAT-AP are broad and inclusive, aiming to embrace data types arising from diverse communities. Therefore, we view the catalogue component of a vocabulary hub as a DCAT Catalogue that hosts DCAT Datasets, representing vocabularies which can be considered as datasets. These vocabularies are available for access or download in one or more formats in our vocabulary repository.
DCAT-AP standardises information on dataset attributes, including descriptions, publishers, and version control. It allows a vocabulary to be any specification, from spreadsheets, ontologies, and JSON Schema to specialized formats. DCAT-AP does not assume anything

about the vocabulary specification format but distinguishes between its various distributions. Therefore, DCAT-AP can be used to specify metadata on the vocabulary with reference to one or more distributions of the vocabulary in a vocabulary repository. Figure 2 illustrates this idea by applying it to the example outlined in chapter 3. This figure shows how to use DCAT-AP to describe the metadata of vocabularies.

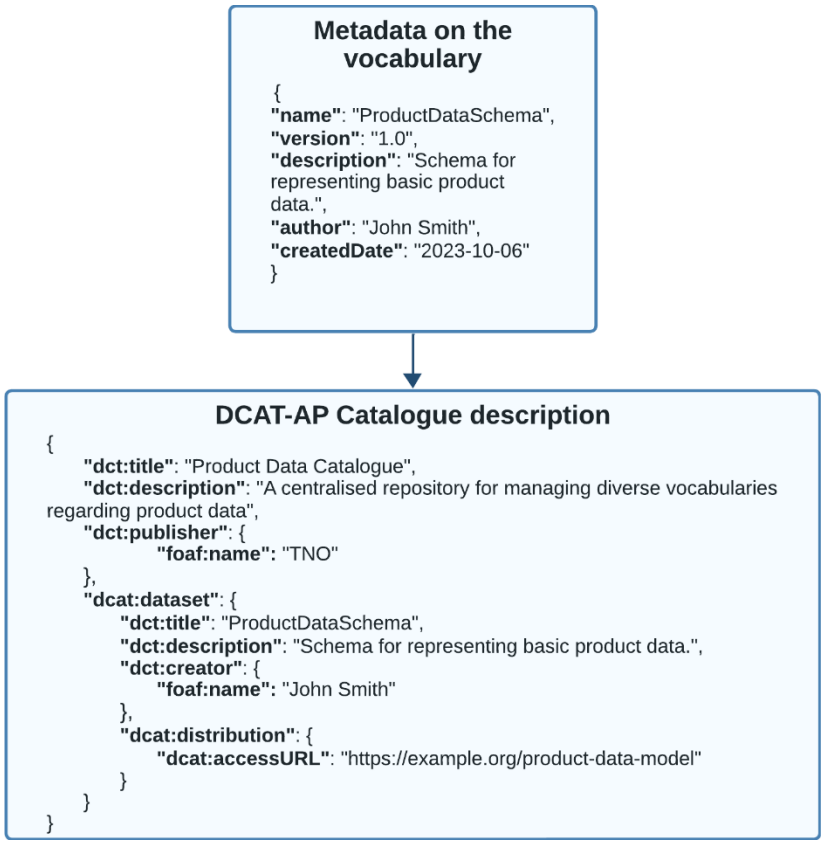

Figure 2: DCAT-AP Catalogue description including a vocabulary as DCAT Dataset

4.2.2 Sharing vocabularies using DCAT-AP

Once a vocabulary and a vocabulary hub are described using DCAT-AP, this representation can be used in the federation of data spaces. Figure 3Error! Reference source not found. shows that the example in Error! Reference source not found.Figure 2, described using DCAT-AP, is being exchanged between two vocabulary hubs. The data shared contains "dcat:accessURL" that specifies how to access the distribution of the actual vocabulary. This can be an URL that enables retrieving a vocabulary via the API of the vocabulary repository in a vocabulary hub.

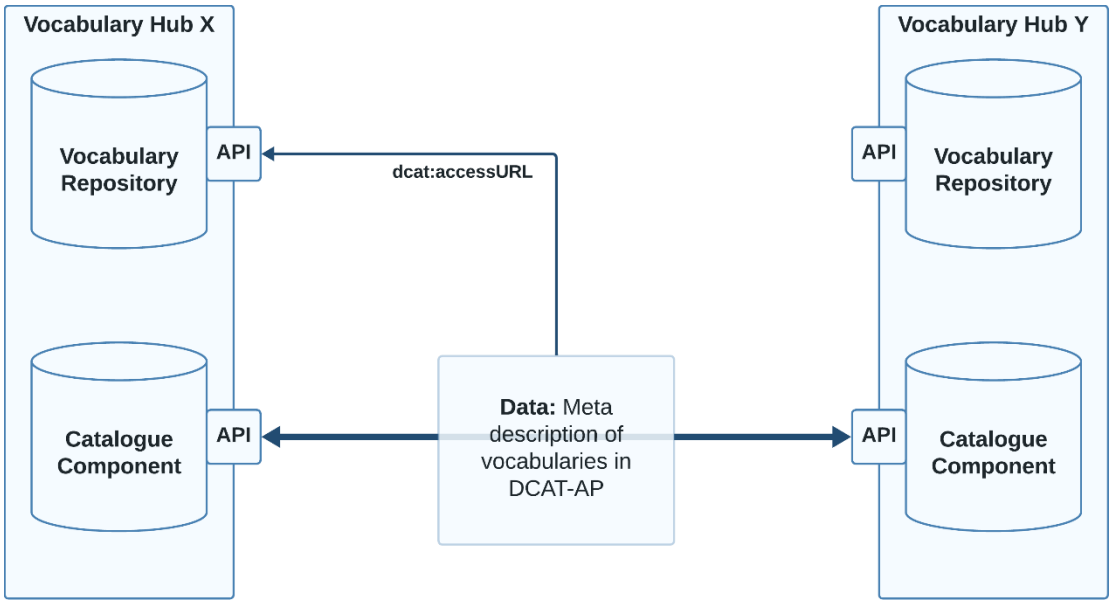

Figure 3: Exchanging vocabularies using DCAT-AP

5 Proof of Concept

A proof of concept has been built to demonstrate how our vision to describe and export vocabularies in DCAT-AP can be operationalized. In the proof of concept, the vocabulary hub Semantic Treehouse has been utilized as the test environment. This chapter describes details about the current implementation, shows example output, validation details and describes final work to be done.

5.1 Mapping to DCAT-AP

This section details the mapping from the current structure describing vocabularies in Semantic Treehouse to the DCAT-AP standard version 2.1.0.
As introduced earlier, a vocabulary is any specification that can be used to consistently describe a set of concepts or data. In Semantic Treehouse, all specifications fall into one of six types, such as message model specification (traditionally XSD messages) or ontologies. Multiple versions can exist for each specification, representing different releases or iterations of the same specification. These specification types and their versions can be distributed in seven types of export formats, including JSON Schema and OpenAPI specification. Overall, an implementation of Semantic Treehouse is called an implementation of a vocabulary hub and contains one or more projects that serve as organized catalogues for related specification and their versions.
We choose to map the Semantic Treehouse structure to DCAT-AP version 2.1.0. as depicted in the following table:

Table 2: Mapping from Semantic Treehouse structure to DCAT-AP v2.1.0.

Semantic Treehouse structure	DCAT-AP
Project	DCAT Catalogue
Specification	DCAT Dataset
Version of a specification	DCAT Dataset
Distribution of a specification	DCAT Distributions

5.1.1 Example of DCAT-AP export output

For the proof of concept export functionality is built into Semantic Treehouse, which makes it possible to export each project to DCAT-AP, and to export the entire vocabulary hub with all projects to DCAT-AP. Two export options are currently provided: a button triggering the browser to download the content in DCAT-AP, and an API is available for machine-to-machine implementation, offering the same export functionality.
The export output is a ttl (Turtle) file that contains all specifications, their versions, and references to their distributions. Within each project catalogue, all the specifications are bundled in a DCAT Catalogue. When exporting the content of the entire vocabulary hub, it is represented as a DCAT Catalogue, including all project catalogues expressed as DCAT Catalogues.

The screenshot below illustrates an exported project containing a specification version named Human Resource Message version 1.3.1. This example represents a vocabulary used in the flexible staffing industry, where the Human Resource Message is used to match a human resource to an open [11]. The first part provides metadata about the specification, which facilitates discoverability by providing sufficient information to grasp the subject of a specification. In the second part, all distributions of the specification version are included. In this case, an XML schema for the Human Resource Message version 1.3.1 is included, with references to access or download the distribution of a specification.

```
<https://setu.staging.semantic-treehouse.nl/specversions/MessageModelVersion_35478627-b2c
  a dcat:Dataset ;
  dc:description "The SETU standard for Ordering and Selection is used for matching a hum
  dc:title "SETU HumanResource v1.3.1"@en ;
  dcat:distribution <https://setu.staging.semantic-treehouse.nl/specversions/MessageModel
  dc:publisher <https://setu.staging.semantic-treehouse.nl/groups/SETU> ;
  dc:identifier "https://setu.staging.semantic-treehouse.nl/specversions/MessageModelVers
  dcat:landingPage <https://setu.staging.semantic-treehouse.nl/#/Message_32_model/Message
  dc:issued "2015-06-16"^^xsd:date ;
  owl:versionInfo "1.3.1 (RELEASE)" .

<https://setu.staging.semantic-treehouse.nl/specversions/MessageModelVersion_35478627-b2c
  a dcat:Distribution ;
  dcat:accessURL <https://setu.staging.semantic-treehouse.nl/api/v1/fit/message/Property_
  dc:description "Distribution of type XSD for STH specification version with id MessageM
  dc:format <http://publications.europa.eu/resource/authority/file-type/SCHEMA_XML> ;
  dcat:downloadURL <https://setu.staging.semantic-treehouse.nl/api/v1/fit/message/Propert
  dcat:mediaType <https://www.iana.org/assignments/media-types/application/xml> ;
  dc:title "XSD schema distribution for MessageModelVersion_35478627-b2d0-4bce-baec-5776b
```

Figure 4: A cropped fragment of an export of the “Human Resource” message model as a DCAT Dataset and an XSD distribution.

5.1.2 Validation

As a means of testing our proof of concept, a generated DCAT-AP Catalogue containing all types of specifications were exported to a file and that file was uploaded to a European validation service for DCAT-AP [14].The generated code passed the most important validity checks. The limited checks that failed had to do with the specific requirements of the validation service that focuses on a dialect of DCAT-AP with additional business rules, like the rule: “Catalogue Publishers need to be from a list of recognized authorities”.

6 Conclusion

Every data space develops its own vocabularies and vocabulary hubs. A standardized solution is necessary to exchange vocabularies, bridging the gap in semantic interoperability across various data spaces. This paper has illustrated that vocabularies can be considered as datasets, enabling them to be described and shared using DCAT-AP. This serves as the starting point towards achieving semantic interoperability, as now vocabularies can be shared across data spaces to encourage maximal reuse of vocabularies across sectors.
The proof of concept has proven the operational feasibility of describing and exporting vocabularies in DCAT-AP. In this process, the Semantic Treehouse vocabulary hub served as the testing environment. This standardized approach allows data space participants to explore vocabularies from other hubs as if they were present in their own. This facilitates the creation of negotiating on semantics across data spaces, reducing the need for each data space to independently create and maintain its own. In essence, this marks the beginning of fostering semantic interoperability in federated data spaces.

6.1 Future work

As we conclude the current phase of our research and implementation, it is essential to outline potential directions for future work. The proposed approach currently focuses solely on DCAT-AP; however, considerations of other standards should be explored in the future. The work so far has laid the foundation for sharing vocabularies across data spaces, marking the starting point for maximizing vocabulary reuse or negotiating a common vocabulary. Some data space specifications already address contractual negotiations within data spaces, such as the data space protocol defining the contract negotiation protocol. In the future, attention should be paid towards determining the feasibility of agreeing upon semantics using this protocol or other specifications.
To facilitate the sharing of vocabularies between vocabulary hubs and other data space components (e.g., connectors and metadata brokers), it will be necessary to develop importing functionality for DCAT-AP representations of external vocabularies. Additionally, addressing access control over non-public specifications is crucial, considering that some vocabularies are not freely downloadable.
Another aspect to address is the establishment of conditions and criteria for vocabularies intended for exchange. Vocabularies should meet specific criteria, including well-written documentation and permanent accessibility through a permanent URL.

7 References

[1] European Commission, “A European strategy for data,” 2020. [Online]. Available: https://digital-strategy.ec.europa.eu/en/policies/strategy-data.

[2] European Commission , “DCAT Application Profile for data portals in Europe,” [Online]. Available: https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/dcat-application-profile-data-portals-europe/release/200.

[3] TNO, “Semantic Treehouse,” TNO, [Online]. Available: https://www.semantic-treehouse.nl. [Accessed 1 11 2023].

[4] International Data Spaces , “Meaningful data,” [Online]. Available: https://docs.internationaldataspaces.org/ids-knowledgebase/v/ids-ram-4/context-of-the-international-data-spaces/2_1_data-driven-business_ecosystems/2_5_meaningful_data.

[5] International Data Spaces, “Vocabulary Hub - IDS Knowledge Base,” [Online]. Available: https://docs.internationaldataspaces.org/ids-knowledgebase/v/ids-ram-4/layers-of-the-reference-architecture-model/3-layers-of-the-reference-architecture-model/3_5_0_system_layer/3_5_6_vocabulary_hub.

[6] European Union, “New European Interoperability Framework (EIF) – Promoting seamless ervices and data flows for European public administrations,” 2017. [Online]. Available: https://ec.europa.eu/isa2/sites/default/files/eif_brochure_final.pdf.

[7] Catena-X, “Catena-X Your Automotive Network,” [Online]. Available: https://catena-x.net/en/.

[8] Smart Connected Supplier Network, “Smart Connected Supplier Network,” [Online]. Available: https://smart-connected.nl/nl.

[9] Smart Data Models, [Online]. Available: https://smartdatamodels.org/.

[10] Ketenstandaard Bouw en Techniek, “Ketenstandaard,” [Online]. Available: https://ketenstandaard.nl/.

[11] SETU, “SETU standaarden voor flexible arbeid,” [Online]. Available: https://setu.nl/.

[12] W3C, “Data Catalog Vocabulary (DCAT) - version 3,” 07 March 2023. [Online]. Available: https://www.w3.org/TR/vocab-dcat-3/.

[13] International Data Spaces, “Governance for Vocabularies,” [Online]. Available: https://docs.internationaldataspaces.org/ids-knowledgebase/v/ids-ram-4/perspectives-of-the-reference-architecture-model/4_perspectives/4_3_governance_perspective/4_3_12_goverance_for_vocabularies.

[14] EC DG DIGIT Interoperability Test Bed, “BRegDCAT-AP validator,” EC DG DIGIT, [Online]. Available: https://www.itb.ec.europa.eu/shacl/bregdcat-ap/upload. [Accessed 1 11 2023].

ICT, Strategy & Policy

www.tno.nl

TNO innovation for life