Report on
OntoCommons ontology registry infrastructure

Grant Agreement: 958371

OntoCommons - Ontology-driven data documentation for Industry Commons, has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement no. 958371.

Project Title	Ontology-driven data documentation for Industry Commons
Project Acronym	OntoCommons
Project Number	958371
Type of project	CSA - Coordination and support action
Topics	DT-NMBP-39-2020 - Towards Standardised Documentation of Data through taxonomies and ontologies (CSA)
Starting date of Project	01 November 2020
Duration of the project	36 months
Website	www.ontocommons.eu

Report on OntoCommons ontology registry infrastructure

Work Package	Ontology Commons Ecosystem Toolkit
Task	Reference implementation tooling
Lead author	María Poveda-Villalón (UPM)
Contributors	Andrea Cimmino (UPM), Raúl García Castro (UPM)
Peer reviewers	Arkopaul Sarkar (ENIT), Lan Yang (NUIG)
Version	Final
Date	04/02/2022

Glossary of terms

Item	Description
RDF	Resource Description Framework
OWL	Web Ontology Language
SPARQL	SPARQL Protocol and RDF Query Language
URI	Uniform Resource Identifier

Keywords

Ontology; Ontology catalogue, Web interface

Disclaimer

OntoCommons.eu has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement no. 958371. The content of this document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content. The European Commission is not liable for any use that may be made of the information contained herein.

Executive Summary

This document describes the workflow followed to generate the "Report on OntoCommons ontology catalogue" and the technological infrastructure developed and deployed to support the creation of this catalogue. In this deliverable some process details are described for example for gathering and cleaning data as well for the transformation into RDF. In addition, SPARQL queries defined to support the portal are provided. Finally, an overview of the generated portal is included. It should be noted that the infrastructure deployed for the catalogue could be reused and extended for other project metadata publication in RDF in the future.

Table of Contents

List of Figures

List of Tables

1. Introduction

On the one hand, one of the OntoCommons goals is to identify and review semantic resources related to main OntoCommons related domains such as industry, manufacturing, construction, and materials, etc. On the other hand the task "Reference implementation tooling" is devoted to the more technical aspects of the ontology development support. More precisely, this report is intended to provide the technical details and description of the infrastructure of the OntoCommons ontology catalogue. That is, this report represents the technical foundation for the ontology.

In this sense, the OntoCommons ontology catalogue is provided as a website available at https://data.ontocommons.linkeddata.es/index and the present document describes the process followed to generate and publish the catalogue, the technologies involved, an overview of the resulting catalogue, and the future lines of work to come. More precisely, the catalogue HTML views are generated dynamically from RDF [Brickley, 2004] data describing the metadata ontologies. That is, the catalogue is built on top of semantically annotated data and is also provided as knowledge graph by means of an SPARQL endpoint available for external queries. For automatically generating the HTML views the Helio framework[1] in combination with Thymeleaf[2], as server-side Java template engine, is used. Helio is a framework that allows the generation and publication as Linked Data service of RDF data from different heterogeneous sources.

The workflow followed to generate the catalogue is detailed in Section 2, while an overview of the resulting catalogue is provided in Section 3. Finally, Section 4 provides some future lines of work.

As specified by the proposal, this effort is geared towards building a tangible, usable, and maintained software system to provide prospective adopter of ontologies an integrated access to existing and newly built ontologies as part of the harmonized ontology ecosystem. In that context, OntoCommons ontology catalogue encodes machine readable data for these semantic assets only at the meta-level. At the same time, OntoCommons ecosystem will also endorse one or more ontology repositories (e.g., IndustryPortal[3], MatPortal[4]) for persistent storage of the ontology source files along with version management, FAIR metadata, evaluation metrics, and content browsing services, including vocabulary search, annotator, recommender and mapping, based on NCBO BioPortal technology[5]. OntoCommons ontology catalogue will align its entries with the corresponding sources from the aforementioned repositories. The strategy for such alignment is given in Section 20

[1] https://oeg-upm.github.io/helio/
[2] https://www.thymeleaf.org/
[3] https://industryportal.enit.fr/
[4] https://matportal.org/
[5] https://ontoportal.org/the-ontoportal-virtual-appliance/ (developed and maintained by Standford University)

2. OntoCommons ontology catalogue generation

This section is devoted to the description of the workflow followed to generate the ontology catalogue. An overview of the steps followed, intermediate resources generated and technologies used will be provided.

As already mentioned, the ontology metadata that will populate the ontology registry has been gathered in close collaboration with project partners. In order to collect ontologies and their metadata in the domains of interest, an online survey was set up. This survey was completed by community members external to the OntoCommons project and by project partners. Such survey has served as input to the ontology landscape analysis reported "Report on existing domain ontologies in identified domains" [Le Franc et al., 2021]. For more information about the ontologies gathered, we advise readers to consult this report.

The process followed to generate the ontology catalogue is depicted in Figure 1 where white boxes with rounded corners represent steps or processes, purple (or shadow if printed in black and white) symbols represent generated resources (if the symbol is a document shape) or generated databases (if the symbol is a database shape). Solid arrows between steps indicate the workflow order, and dashed arrows indicate inputs to and outputs from the different steps.

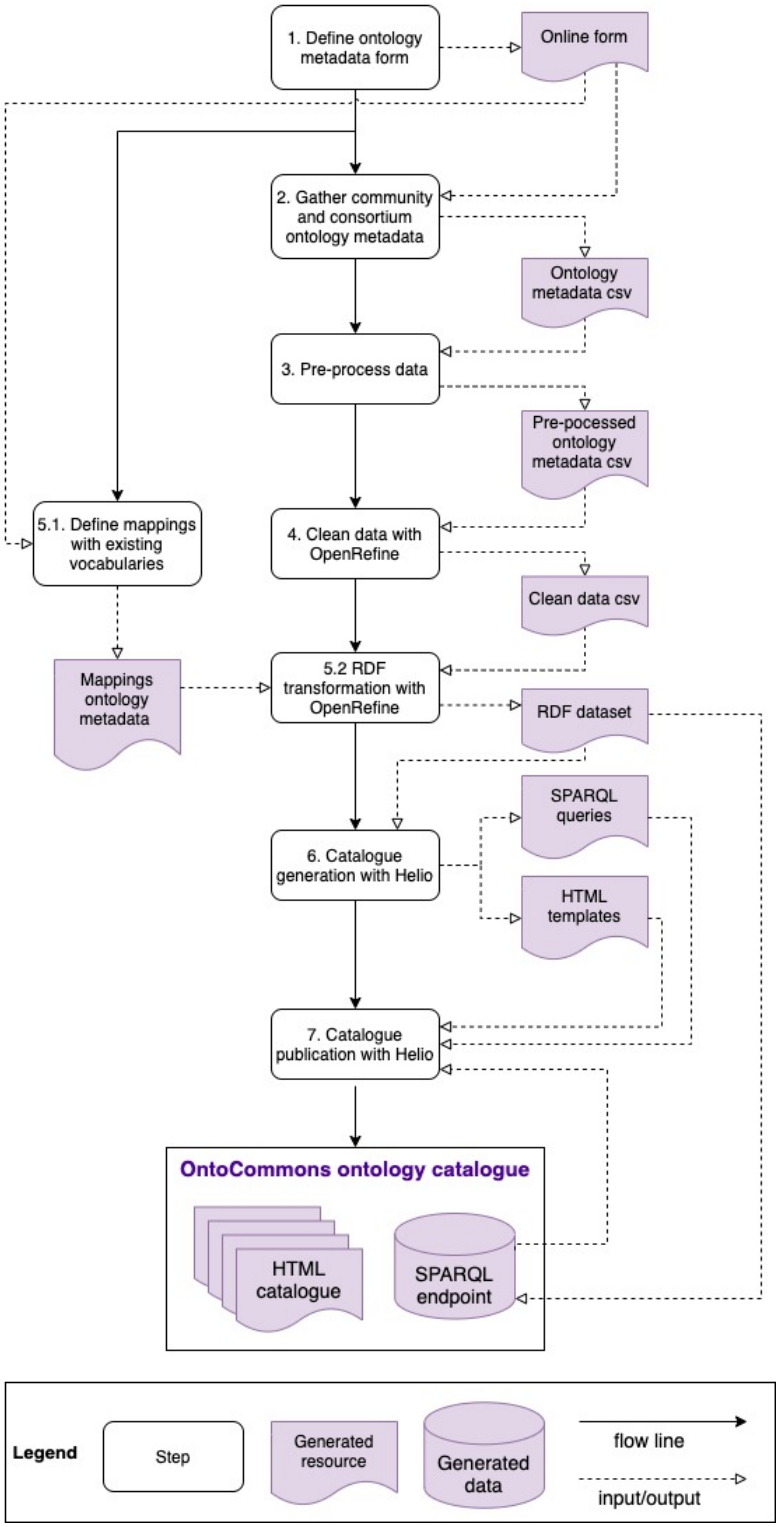

Figure 1. OntoCommons ontology catalogue generation workflow

In the following each step is explained in more detail:

- **Step 1: Define ontology metadata form**
  - **Input**: existing forms from ontology surveys
  - **Process**: the project partners held meetings and discussions to agree on the survey questions and the intended metadata to be requested. The survey was reviewed by the partners and when a decision was made the survey was published at https://ontocommons.eu/node/146.
  - **Output**: Survey online
- **Step 2: Gather community and consortium ontology metadata**
  - **Input**: Survey online
  - **Process**: In order to involve different stakeholders or communities members to provide metadata about ontologies, the survey has been promoted in social media and other channels. In addition, a list of ontologies was elaborated by OntoComons partners, and responsible organizations were assigned to provide the metadata for a number of ontologies.
  - **Output**: CSV file with 45 answers (note: this data represent the answers at the moment of starting the ontology catalogue generation and that duplicates are counted individually).
- **Step 3: Pre-process data**
  - **Input**: CSV file with 45 answers
  - **Process**: During this step, a pre-cleaning process was carried out to discard invalid inputs. It should be clarified that this process is different from the data cleaning in the next step. The goal of this step is to remove invalid rows rather than clean or correctly format valid data. Some examples of actions taken during this are:
    - Remove duplicates for MAMBO entry.
    - Remove the link to another survey, as this entry was informative and no ontology was reported.
    - Remove VIMMP ontologies as it points to a zenodo entry and it represents a set of ontology and each entry should be one ontology. Responsible persons will be emailed to solve this situation and include one entry per ontology.
    - Remove the FLEXINET ontology for Product-Service Production, as the URI provided is an URL for a pdf instead of a valid URI. The responsible persons will be emailed to solve this situation.
    - Remove the Dicon ontology network as each entry should be one ontology, not a portal. In addition, some of the ontologies from this network are already included as entries in the survey.
    - Remove the Supply Chain ONTOlogy (SCONTO) as each entry should be one ontology. Responsible persons will be emailed to solve this situation and include one entry per ontology.
  - **Output**: Pre-processed ontology metadata CSV containing 38 entries

- **Step 4: Clean data with OpenRefine**
  - **Input**: Pre-processed ontology metadata CSV
  - **Process**: During this process, the data has been formatted, clean, and prepared to ease the transformation to RDF. The main operations carried out during these steps have been:
    - Split multivalued cells into several rows, for example, the "domain" field.
    - Format "scope" filed due to the difference styles used in the raw data.
    - Replace the natural languag codes used in the raw data with the ISO-369-3 codes to link to the Lexvo datasets.
    - Replace ontology languages and formats with Wikidata entities in order to link to the Wikidata dataset.
    - Remove "other" values when the meaning is to indicate that the value is provided in the next column as the option was not provided in the survey.
  - **Output**: Clean CSV data.
- **Step 5.1: Define mappings with existing vocabularies**
  - **Input**: Survey online
  - **Process**: In order to generate the semantically annotated data in RDF the ontologies and the elements to describe such data should be defined. For doing so, the OntoCommons partners reviewed the online survey form to match each question with existing ontologies. The prefixes and namespaces of the reused ontologies are listed in Table 1 and the agreed correspondences between the survey fields and the different metadata ontologies analyzed are provided in Table 2.
  - **Output**: Mappings between the survey fields and metadata ontologies elements. These mappings are provided in Table 2.

Table 1. Vocabularies reused to describe ontologies metadata within the OntoCommons catalogue

Vocabulary	Prefix	URI
Creative Commons Rights Expression Language	cc	http://creativecommons.org/ns
Dublin Core Metadata Initiative Metadata Terms	dc	http://purl.org/dc/terms/
Friend of a Friend	foaf	http://xmlns.com/foaf/0.1/
MOD: Metadata for Ontology Description and publication [Biswanath et al., 2015]	mod	https://w3id.org/mod#
Ontology Metadata Vocabulary [Hartmann et al., 2005]	omv	http://omv.ontoware.org/2005/05/ontology#
Web Ontology Language	owl	http://www.w3.org/2002/07/owl#
VANN: A vocabulary for annotating vocabulary descriptions	vann	http://purl.org/vocab/vann/
Vocabulary of a Friend	voaf	http://purl.org/vocommons/voaf#

Table 2. Mappings between survey fields and existing ontology properties

Question	Description	Mapping
Name (mandatory)	The name given to the ontology.	dc:title
URI (mandatory)	The URI of the ontology.	vann:preferredNamespaceUri
Description (mandatory)	A free text account of the ontology.	dc:description
Domains (mandatory)	The different domains covered by the ontology. If the ontology covers more than one domain, please separate them by commas. Example: manufacturing, material science, maintenance, AEC industry, marketing, ...	dc:subject
Scope	The scope of the ontology in a particular domain e.g., predictive maintenance, stakeholder description, product nomenclature, sensor, building.	*Column not included in this version due to different uses in the data. It mixes the domains and the description fields.*
Namespace	The preferred namespace URI to use when using terms from this vocabulary.	vann:preferredNamespacePrefix:
Version	The version of the ontology.	owl:versionInfo
Creation date	The date of formal issuance of the ontology.	dc:created
Last update	Most recent date on which the ontology was changed, updated, or modified.	dc:modified
Contact person	The person(s) primarily responsible for creating the ontology. Please include the name and email address of the contact persons whenever possible. If there is more than one contact person, please separate them by commas.	*Field not included in the RDF data and website due to potential privacy issues. For now it is used for internal purposes.*
Publisher	The organization that published the ontology.	dc:publisher
Ontology language	The ontology language in which the ontology is implemented.	omv:hasOntologyLanguage
Format	Format in which the ontology code is provided.	omv:hasOntologySyntax
Use of Top-Level ontologies?	Top level ontologies used by the ontology.	voaf:reliesOn

License (mandatory)	The license of the ontology. Example: CC BY-SA, MIT, etc.	cc:license
Please specify (license)	Specify the license if it is not one of the list.	cc:license
Language	The ISO 639-1 code(s) of the language(s) of the resource. If the ontology is implemented in more than one language, please separate them by commas. Example: es, en, (See http://en.wikipedia.org/wiki/List_of_ISO_639-1_codes for a full list of codes).	dc:language
Available documentation (mandatory)	URLs for the documentation of the ontology (for example a website).	foaf:page
References	Resources that might provide additional information (documents, deliverables, papers, etc.).	dct:bibliographicCitation
Ontology registered	Is the ontology stored and indexed in a dedicated repository/registry? If yes, could you please specify which one and provide the URL of the repository/registry?	dc:isPartOf
Best practices	Free text	
Development methodology and knowledge sources	Please provide a short description of the methodology and knowledge sources used to develop the ontology as a comma separated list.	omv:usedOntologyEngineeringMethodology
Is the ontology an outcome of a European project? If so, please indicate the project name and the website if possible.	Whether the ontology has been developed in one or more European projects.	*Column not included in this version due to different uses in the data.*
Is the ontology developed within a standardization body? If yes, please specify which one	Whether the ontology has been developed in the context of standardization bodies.	dc:contributor
Is the ontology based on any standards? If yes, please specify which one(s)	Whether the ontology is based on existing standards.	dc:conformsTo
Is the ontology supported by a community? If yes, please mention the involved community(ies)	Whether the ontology is being supported by any community.	*Field not included in the RDF data. For now, it is used for internal purposes.*

Is there a sustainability plan for this ontology?	Whether there is a sustainability plan form an organization, community, company, etc.	*Field not included in the RDF data. For now, it is used for internal purposes.*
Is the ontology being reused by other ontologies or projects? If yes, could please specify which ones?	Whether the ontology is being adopted.	voaf:usedBy
Is the ontology aligned with other ontologies, reuse other ontologies or specific design patterns? If yes, please specify which one(s).	Whether the ontology reuses ontology design patterns.	voaf:reliesOn
Comments	Further information about the ontology that might be relevant.	*Field used for internal purposes.*

- **Step 5.2: RDF transformation with OpenRefine**
  - **Input**: Mappings between the survey fields and metadata ontologies elements and CSV data.
  - **Process**: During this step, the mappings defined in Table 2 are encoded into OpenRefine as depicted in Figure 2. In addition, the entities URI strategy should be defined. In particular, in this case the base URI for the generated data is https://data.ontocommons.linkeddata.es/ and for each type of resource a path is defined, namely:
    - For entities of type vocabulary: https://data.ontocommons.linkeddata.es/vocabulary/. An example of vocabulary URI would be https://data.ontocommons.linkeddata.es/Mambo as the identifiers used for generating URIs for vocabularies are based on the title field.
    - For entities of type domain: https://data.ontocommons.linkeddata.es/domain/, an example of domain would be https://data.ontocommons.linkeddata.es/Construction.
  - **Output**: RDF dataset in turtle format[6].

[6] https://www.w3.org/TR/turtle/

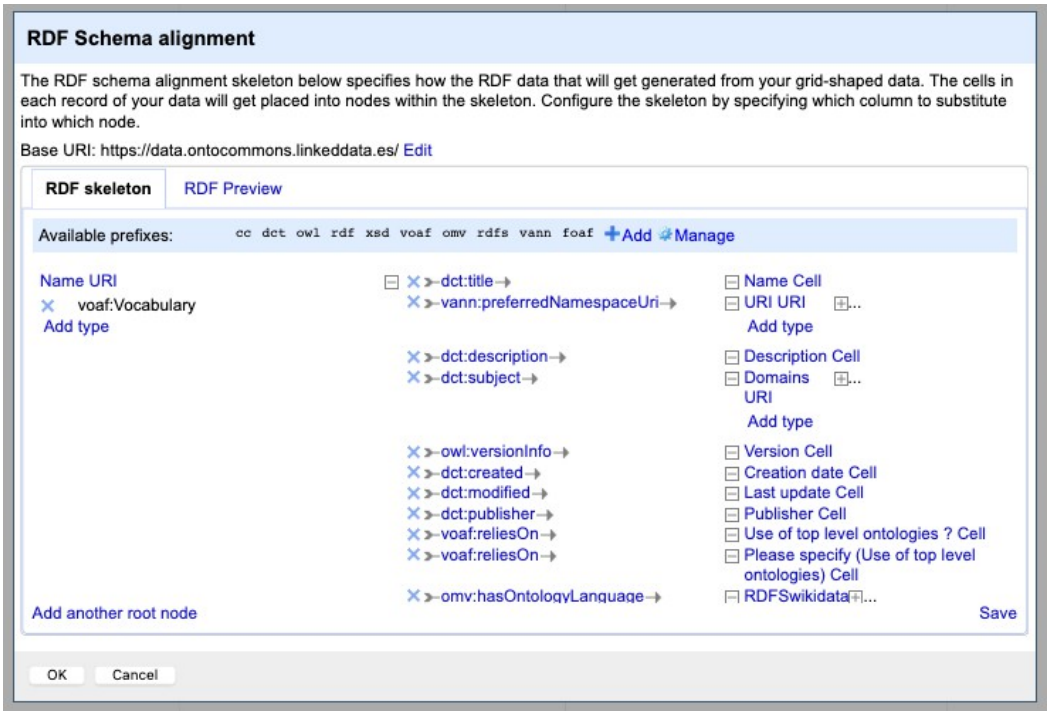

Figure 2. Mappings definition in OpenRefine

- **Step 6: Catalogue generation with Helio**
  - **Input**: RDF dataset
  - **Process**: Taking into account the RDF dataset, during this step the HTML templates and the SPARQL queries needed to visualize the RDF data in a web browser are defined. More precisely, the following views have been defined:
    - Main view or landing page: For the main page of the catalogue, a table showing the ontology title (linking to the vocabulary page within the catalogue), the link to its URI (if available), the license (if available), the ontology language and the syntaxes in which the ontology is provided, the ontology domains and the languages in which the ontology is documented is shown. For generating this view, the SPARQL query shown in Listing 1 has been defined.
    - Vocabulary page: For generating vocabulary pages there is no need for a specific SPARQL query as the information shown is the triples in which the vocabulary acts as a subject; therefore, it is equivalent to a DESCRIBE query
    - Domain page: For each domain, a view showing the ontologies for such domain is generated. For generating this view, the SPARQL query shown in Listing 2 has been defined.
  - **Output**: From this activity, the HTML templates and the SPARQL queries are generated.

Listing 1. SPARQL query for landing page

```
PREFIX dct: <http://purl.org/dc/terms/>
PREFIX voaf: <http://purl.org/vocommons/voaf#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX cc: <http://creativecommons.org/ns#>
PREFIX vann: <http://purl.org/vocab/vann/>
PREFIX omv: <http://omv.ontoware.org/2005/05/ontology#>

SELECT DISTINCT ?sub ?title
(GROUP_CONCAT(DISTINCT ?uri ; separator=' ') AS ?uris)
(GROUP_CONCAT(DISTINCT ?license ; separator=' ') AS ?licenses)
(GROUP_CONCAT(DISTINCT ?domain ; separator=' ') AS ?domains)
(GROUP_CONCAT(DISTINCT ?lang ; separator=' ') AS ?langs)
(GROUP_CONCAT(DISTINCT ?ontlang ; separator=' ') AS ?ontlangs)
(GROUP_CONCAT(DISTINCT ?ontsyntax ; separator=' ') AS ?ontsyntaxs)
WHERE {

   ?sub a voaf:Vocabulary .
   ?sub dct:title ?title .
   OPTIONAL { ?sub vann:preferredNamespaceUri ?uri . }
   OPTIONAL { ?sub cc:license ?license . }
   OPTIONAL { ?sub omv:hasOntologyLanguage ?ontlang . }
   OPTIONAL { ?sub omv:hasOntologySyntax ?ontsyntax . }
   OPTIONAL { ?sub dct:subject ?domain . }
   ?sub dct:language ?lang .
}

GROUP BY ?title ?sub Order by asc(?title)
```

Listing 2. SPARQL query for domain resources

```
PREFIX dct: <http://purl.org/dc/terms/>
PREFIX voaf: <http://purl.org/vocommons/voaf#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX cc: <http://creativecommons.org/ns#>
PREFIX vann: <http://purl.org/vocab/vann/>
PREFIX omv: <http://omv.ontoware.org/2005/05/ontology#>

SELECT DISTINCT ?onto ?title ?sub
WHERE {

   ?onto dct:subject ?sub .
   ?onto a voaf:Vocabulary .
   ?onto dct:title ?title .
}

 Order by asc(?title)
```

- **Step 7: Catalogue publication with Helio**
  - **Input**: RDF data set, HTML templates and SPARQL queries
  - **Process**: This step is dedicated to the setup of the web server and Helio publisher deployment.
  - **Output**: Online website for OntoCommons ontology catalogue including HTML on demand generation and SPARQL endpoint.

3. OntoCommons ontology catalogue overview

The OntoCommons ontology catalogue is provided as a website that automatically generates the different views (main page, vocabularies information, or domain information) on-demand when an URL is requested. The catalogue architecture is depicted in Figure 3. In such figure we can observe that the catalogue is published by Helio Publisher. More precisely, for generating the HTML views, Helio populates the HTML templates with the data retrieved from the SPARQL endpoint using the SARPQL templates. In order to provide the SPARQL endpoint to external users, Helio connects the web user interface with the backend SPARQL endpoint. In addition, it is possible to retrieve the catalogue resource information for vocabularies and domains in RDF by means of content negotiation provided by Helio.

In the particular case of the OntoCommons catalogue, the SPARQL endpoint is deployed over a GraphDB instance, however, other triple stores could be used.

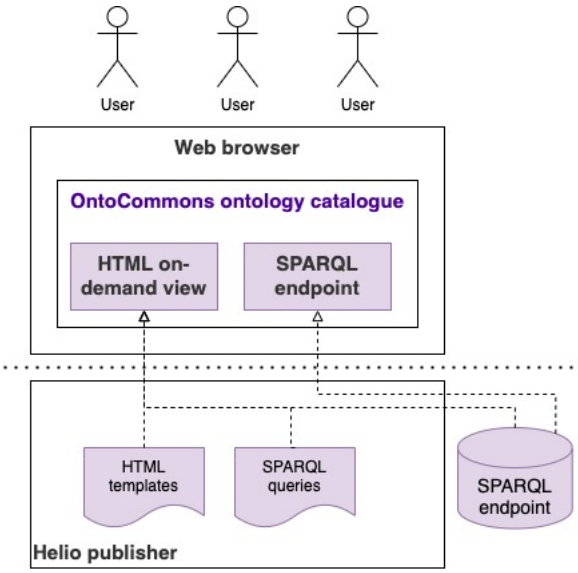

Figure 3. OntoCommons ontology catalogue architecture

In order to generate dynamically the HTML visualizations, the HTML templates are based on the Thymeleaf framework that allows us to process the SPARQL queries results. Some examples of the

pages created are shown below. More precisely, Figure 4 depicts an excerpt of the main or landing page of the catalogue, an example of a generated page for a vocabulary is shown in Figure 5, and finally Figure 6 depicts the information shown for each domain. The mappings between the ontology properties used in the RDF data and the HTML fields for the vocabulary view are shown in

Table 3.

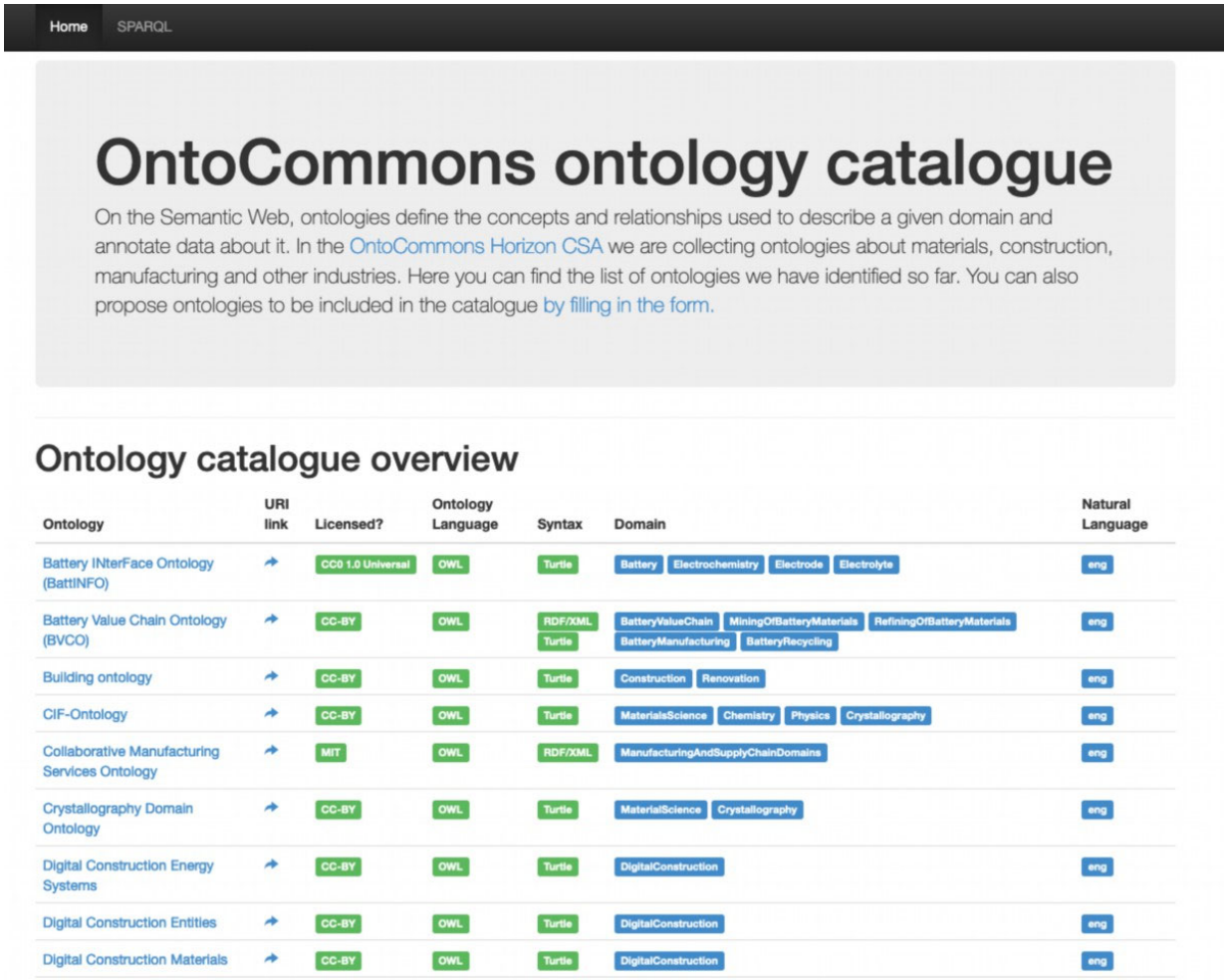

Figure 4. Excerpt from the OntoCommons ontology catalogue landing page

Table 3. Mappings between ontology properties and HTML fields

Ontology property	HTML field
dc:title	Title
vann:preferredNamespaceUri	URI
dc:description	Description
dc:subject	Domains

owl:versionInfo	Version
dc:created	Creation date
dc:modified	Modification date
dc:publisher	Publisher
omv:hasOntologyLanguage	Ontology languages
omv:hasOntologySyntax	Ontology format
voaf:reliesOn	Reused vocabularies
cc:license	License
dc:language	Languages
foaf:page	Website
dct:bibliographicCitation	Bibliographic refs
dc:isPartOf	Registered in
omv:usedOntologyEngineeringMethodology	Methodology used
dc:contributor	Supported by
dc:conformsTo	Based on standards
voaf:usedBy	Used by

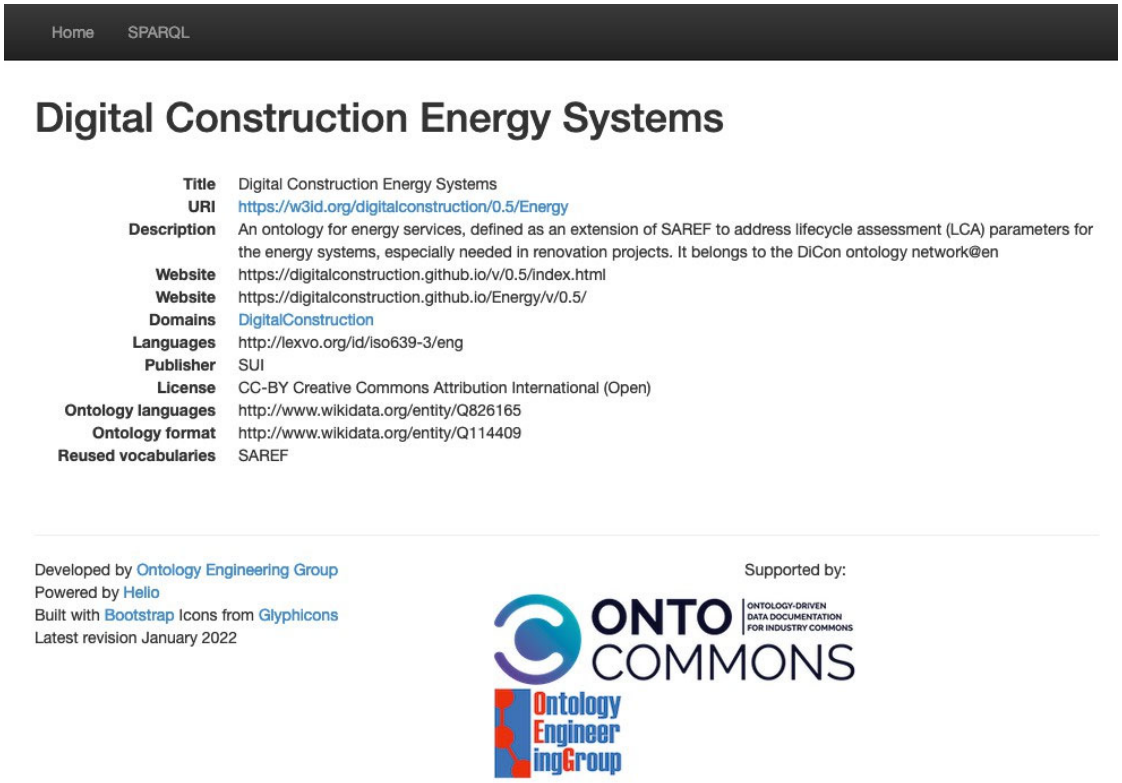

Figure 5. Example of OntoCommons catalogue vocabulary view

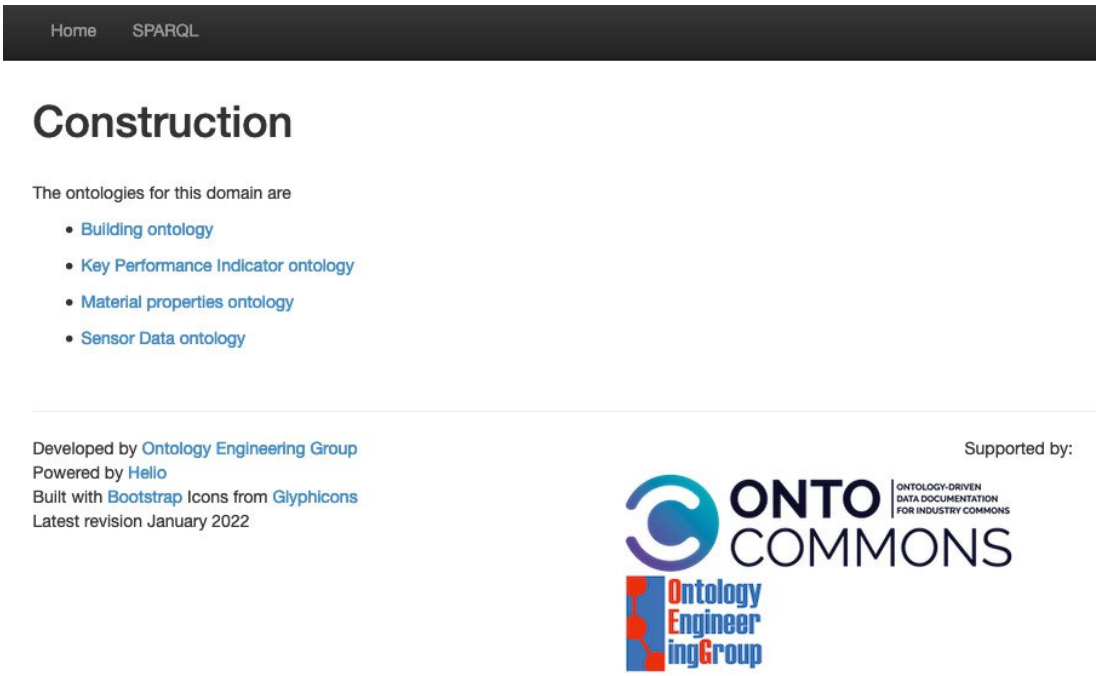

Figure 6. Example of OntoCommons catalogue domain view

4. Conclusions and future work

This document presented the process followed to generate and publish the OntoCommons ontology catalogue based on community input and contributions from project partners.

As the catalogue is intended to evolve over the course of the project, it should not be considered as a final product at the end of the project. Indeed, next action points are already defined:

- Refine input data, for example, to homogenise the information about the project in which the ontologies were developed.
- Contact responsible persons for the ontologies discarded in the preprocess step in order to gather the different input to each ontology instead of ontology networks.
- Include information about which ontologies have been promoted to other existing repositories and registries by OntoCommons partners. For doing this, a new existing ontology property would be selected to distinguish between this case and the information already available about which ontologies are registered in other catalogues. A potential candidate to represent this information is *http://www.w3.org/2006/gen/ont#sameWorkAs* which is described as "The equivalence relation linking all versions of a work, specific or generic along various axes."

Finally, it should be mentioned that due to the architecture deployed based on RDF and customized HTML template in a modular way, the ontology catalogue infrastructure could be used as basis for other project metadata that could be of interest for publication that is represented or intended to be formalized as a knowledge graph. In this sense, for each entity to be represented a new HTML template and possible SPARQL query should be defined.

5. References

[Biswanath et al., 2015] Dutta, Biswanath, Durgesh Nandini, and Gautam Kishore Shahi. "MOD: metadata for ontology description and publication." International Conference on Dublin Core and Metadata Applications. 2015.

[Brickley, 2004] Brickley, D. (2004). *RDF vocabulary description language 1.0: RDF schema*. http://www. w3. org/tr/rdf-schema/.

[Hartmann et al., 2005] Jens Hartmann, Raúl Palma, York Sure, Mari del Carmen Suárez-Figueroa, Peter Haase, Asunción Gómez-Pérez, Rudi Studer. *Ontology Metadata Vocabulary and Applications*. Workshop on Web Semantics (SWWS2005). Agia Napa, Cyprus. 1-2 November 2005.

[Le Franc et al., 2021] Yann Le Franc, Gerhard Goldbeck, Arkopaul Sarkar, Jesper Friis, María Poveda-Villalón, Alba Fernández-Izquierdo, Hedi Karray. "D3.2 Report on existing domain ontoloiges in identified domains". OntoCommons project deliverable. 2021.