Report on existing domain ontologies in identified domains

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 existing domain ontologies in identified domains

Work Package	Industrial Domain Ontologies
Task	Domain-specific semantic Landscape Analysis
Lead author	Yann Le Franc (eSDF)
Contributors	Gerhard Goldbeck (GCL), Arkopaul Sarkar (ENIT), Jesper Friis (SINTEF), María Poveda Villalon (UPM), Alba Fernández Izquierdo (UPM), Hedi Karray (ENIT), Emna Amdouni (ENIT), Emilio Sanfilippo (CNR)
Peer reviewers	Dimitris Kiritsis (UiO), John Breslin (NUIG)
Version	Final
Date	09/03/2022

Glossary of terms

Item	Description
Semantic artefact	Semantic Artefact is defined here as defined in the context of FAIRsFAIR (Le Franc et al., 2020) i.e., as a machine-actionable and -readable formalisation of a conceptualisation enabling sharing and reuse by humans and machines. These artefacts may have a broad range of formalisation, from a loose set of terms, taxonomies, and thesauri to higher-order logics, and include the concepts/terms/classes constituting these artefacts.
Top-Level Ontology	A top-level ontology (or foundation ontology) is an ontology (in the sense used in information science) which consists of very general terms (such as "object", "property", "relation") that are common across all domains[1].
Mid-Level Ontology	A mid-level ontology is one that adds general content to the structure outlined in the upper-level ontology by identifying types of entities which directly specialise the upper-level types, but which are also common to many domains of interest. Classes that appear in mid-level ontologies are still fairly basic with respect to particular knowledge domains and often require further specialisation to be useful for data modeling (e.g., Person, Act of Communication, and Geopolitical Entity)[2].
Domain ontology	A domain-level ontology is one that identifies types that further specialise the basic types from one or more mid-level ontologies. Domain ontologies describe objects, events, and relationships that are of interest to a more limited number of knowledge domains (e.g., Intelligent Analyst Role, Portion of Ammonium Nitrate, or Act of Watercraft Registration)[3].

Keywords

Domain Ontology; Landscape analysis, FAIR evaluation, topological analysis

[1] https://www.nist.gov/system/files/documents/2021/10/14/nist-ai-rfi-cubrc_inc_004.pdf
[2] https://www.nist.gov/system/files/documents/2021/10/14/nist-ai-rfi-cubrc_inc_004.pdf
[3] https://www.nist.gov/system/files/documents/2021/10/14/nist-ai-rfi-cubrc_inc_004.pdf

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

OntoCommons aims at defining a semantic interoperability framework to support the documentation of industrial data with ontologies. This document summarises the landscape analysis on domain ontologies. The scope of this analysis covers the domains of Physics and Chemistry, Mechanical and Industrial Engineering, Materials Science and Engineering, Thermal and Process Engineering, and Computer Science, Systems and Electrical Engineering.

A dataset of 130 ontologies has been created based on expert inputs collected during workshops and surveys. Using this dataset, we collected information both manually and automatically to better describe the landscape (number of ontologies by domains, usage of Top-Level Ontologies, serialisation, complexity, compliance to FAIR principles, domain coverage, etc.).

This first analysis highlighted the strong heterogeneities within and among the different domains and the low level of compliance to FAIR principles for each community.

Table of Contents

List of Figures

List of Tables

1. Introduction

OntoCommons aims at providing an interoperability layer by building a framework for integrating existing domain specific ontologies with Top Level Ontologies (TLO).

To achieve such a goal, it is of paramount importance to investigate and understand the existing landscape of domain ontologies to provide a basis for such integration.

In this report, we describe our attempt to survey the landscape of domain ontologies in the areas of Materials Science and Manufacturing. We first provide some context for the analysis with respect to the considered domains and the “FAIRness” evaluation. We then describe our methodology for collecting ontologies into a dataset and for analysing the dataset itself. The results of the analysis are shown in the following sections. Finally, we draw some conclusions on this work.

1.1 Context

The core objective is to collect community inputs and formulate a framework for harmonising domain ontologies in order to improve intra- and cross-domain interoperability.

In this context, the aim is to identify existing semantic resources developed and potentially used for Materials Science and Manufacturing. Our objective is to provide a qualitative view of the semantic landscape in the different domains and subdomains covered by Materials Science and Manufacturing.

This aggregated view is crucial to identify gaps and needs for the domains of interest and to define a framework for intra- and cross-domain interoperability, as well as a review of domain interoperability.

Our first task was to scope the landscape analysis by defining the domains of interest. The outcome of this work is described in the following section.

As FAIR principles provide the foundations for intra- and cross-domain data interoperability, we discuss ongoing efforts geared toward providing recommendations and tools for evaluating the FAIRness of our data i.e., ontologies, or more broadly semantic artefacts, with FAIRness being defined as the degree to which these digital resources are findable, accessible, interoperable, and reusable.

1.2 Scoping the landscape survey: defining domains of interest

Dealing with the landscape analysis of domain ontologies without a clear scope can be extremely cumbersome and time consuming. To support our effort, we started by defining a classification of existing scientific and engineering domains to organise the ontologies and capture their domain coverage. This domain classification is then used to annotate the ontologies that were collected in the landscape analysis. Such information will be used to facilitate their retrieval within the registry which will be populated with the ontologies collected in the landscape analysis. This approach is used to support retrieval in existing community-specific semantic artefact/ontology repositories like Agroportal[4] (domain categorisation) and Bioportal[5] (subject topic).

At the beginning of the landscape analysis, the domain experts group identified four main domains relevant to the scope of the project, namely:

1. D1: Physics and Chemistry
2. D2: Mechanical and Industrial Engineering
3. D3: Thermal Engineering / Process Engineering
4. D4: Materials Science and Engineering

These high-level domains have been derived from existing classifications of scientific domains such as the classification provided by the German Research Foundation (DFG)[6]. The suitability of this classification for our purposes has been checked by the domain experts involved in the project. The table below shows the main categories reused from the DFG classification. Relevant categories for the purposes of OntoCommons that are not found in the DFG classification have been inherited, whenever possible, from the ERC classification areas[7] provided by the European Commission; in the remaining few cases, we introduced new categories according to the analysis of the initial 11 demonstrators of the project. These categories are marked with a star in the table.

With respect to the D1-D4 categories mentioned above, D2-D4 have a direct match to the DFG categories, whereas D1 merges Chemistry (DFG reference code 31) and Physics (32). To refine our classification of ontologies, we have in addition imported the Research area: Computer Science, Systems and Electrical Engineering (44) which was not explicitly mentioned by the domain experts group, considering the relevance of such fields for the project.

*Table 1 - Domain classification (new areas included in existing classification are marked with *)*

Research area	Research subarea	DFG reference code	ERC reference code
Chemistry (D1)		31	
	Molecular Chemistry	301	
	Chemical Solid State and Surface Research	302	

[4] http://agroportal.lirmm.fr/
[5] https://bioportal.bioontology.org/
[6] The DFG domain classification is available at: https://www.dfg.de/download/pdf/dfg_im_profil/gremien/fachkollegien/amtsperiode_2016_2019/fachsystematik_2016-2019_en_grafik.pdf
[7] The ERC classification areas is available at: https://erc.europa.eu/sites/default/files/document/file/erc%20peer%20review%20evaluation%20panels.pdf

	Physical Chemistry of Solids and Surfaces, Material Characterisation	302-02	
	Physical and Theoretical Chemistry	303	
	Analytical Chemistry, Method Development (Chemistry)	304	
	Biological Chemistry and Food Chemistry	305	
	Polymer Research	306	
Physics (D1)		32	
	Condensed Matter Physics	307	
	Optics, Quantum Optics and Physics of Atoms, Molecules and Plasmas	308	
	Particles, Nuclei and Fields	309	
	Statistical Physics, Soft Matter, Biological Physics, Nonlinear Dynamics	310	
Mechanical and Industrial Engineering (D2)		41	
	Production technology	401	
	Metal-cutting manufacturing engineering	401-01	
	Joining, Mounting and Separation Technology	401-03	
	Plastics Engineering	401-04	
	Production Management and Operations Management	401-05	
	Machine Tools and Production Automation	401-06	
	*Additive manufacturing**		
	Mechanics and constructive mechanical engineering	402	
	Engineering design, machine elements, product development	402-01	
	Mechanics	402-02	
	Lightweight Construction, Textile Technology	402-03	
	Acoustics	402-04	

	Aerospace Engineering		PE8_1
	*Automotive Engineering**		
	*Industrial Maintenance**		
	*Procurement, supplier and vendor engineering**		
	*Quality Control**		
Thermal Engineering/ Process Engineering (D3)		42	
	Process engineering, technical chemistry	403	
	Chemical and Thermal Process Engineering	403-1	
	Technical Chemistry	403-2	
	Mechanical Process Engineering	403-3	
	Biological Process Engineering	403-4	
	Heat energy technology, thermal machines, fluid mechanics	404	
	Energy Process Engineering	404-1	
	Technical Thermodynamics	404-2	
	Fluid Mechanics	404-3	
	Hydraulic and Turbo Engines and Piston Engine	404-4	
Materials Science and Engineering (D4)		43	
	Materials engineering	405	
	Metallurgical and thermal processes, thermomechanical treatment of materials	405-01	
	Sintered metallic and ceramic materials	405-02	
	Composite materials	405-03	
	Mechanical behaviour of construction materials	405-04	
	Coating and Surface Technology	405-05	
	*Industrial bioengineering**		
	*Industrial biofuel production**		

	*Structuring and Functionalisation**		
Computer Science, Systems and Electrical Engineering (D5)		44	
	Systems engineering	407	
	Simulation engineering and modelling		PE7-3
	Networks (communication networks, sensor networks, networks of robots...)		PE7-8
	Automation, control systems, robotics, mechatronics, cyber physical systems	407-01	
	Measurement Systems	407-02	
	Microsystems	407-03	
	Traffic and Transport Systems, Logistics, Intelligent and Automated Traffic	407-04	
	Human Factors, Ergonomics, Human-Machine System	407-05	
	Biomedical Systems Technology	407-06	

The classification has been used to annotate the dataset of ontologies collected during the landscape analysis. Our first annotation is leveraging the high-level domains. In this version of the document, we present a refined classification by adding subdomain(s) for each ontology (see Section 3 - Results).

The next step for the landscape analysis is to investigate the relationship between the existing ontologies and FAIR principles. This is described in the next section.

1.3 Evaluating FAIRness of ontologies

FAIR principles (Wilkinson et al., 2016) are recommendations to build an efficient and machine-friendly data environment. As part of these recommendations, principle I2 ("(Meta)data use vocabularies that follow FAIR principles") requires the usage of FAIR ontologies, controlled vocabularies, and any other semantic artefact (for definition, see Glossary and Le Franc et al., 2020). To support the alignment of semantic artefacts with the FAIR principles, the FAIRsFAIR[8] project developed a set of recommendations. A first community-based version has been released in March 2020 (Le Franc et al., 2020) and a second release incorporating community feedback and aligned with RFC 2119[9] has been published in January 2021 (Hugo et al., 2021). These recommendations are used as a basis for evaluating FAIRness in this work. At the time of writing, FAIRsFAIR has not yet

[8] https://www.fairsfair.eu/

[9] Key words for use in RFCs to Indicate Requirement Levels: https://datatracker.ietf.org/doc/html/rfc2119

proposed an evaluation grid with metrics to quantify the level of FAIRness of ontologies. To pursue our goal, we have analysed the recommendations and established an initial set of indicators to evaluate a subset of ontologies from our initial dataset. This work strengthens the relation between OntoCommons and FAIRsFAIR and supports the objective to link OntoCommons with FAIR- and EOSC-related projects.

In the following section, we will describe the FAIR Semantics recommendations and provide a list of indicators.

1.3.1 *FAIR Semantics recommendations*

The FAIRsFAIR project proposes 17 high-level recommendations and 10 best practices to make semantic artefacts FAIR. In the context of this work, we focused our attention to the high-level recommendations which have been validated by different communities. We willingly do not consider the best practices which still need to be aligned with other existing best practices such as those proposed by Garijo et al. (Garijo et al., 2020).

As shown in the table below, these recommendations can be grouped along four main topics: identifiers, metadata, semantic repositories, and semantic alignment.

Table 2 - FAIR Semantics Recommendations.

Rec #	Recommendation	FAIR principles
Identifier		
P-Rec. 1	Globally Unique, Persistent and Resolvable Identifiers **MUST** be used for Semantic Artefacts, their content (terms/concepts/classes and relations) and their versions	F1
P-Rec. 2	Globally Unique, Persistent and Resolvable Identifiers **MUST** be used for Semantic Artefact Metadata Record. Metadata and data must be published separately, even if it is managed jointly	F1, F3
Metadata		
P-Rec. 3	A common minimum metadata schema **MUST** be used to describe semantic artefacts and their content	F2, R1.1, R1.2 and R1.3
P-Rec. 8	Human and machine-readable persistence policies for semantic artefacts metadata and data **MUST** be published	A2
P-Rec. 9	Semantic artefacts **MUST** be made available as a minimum portfolio of common serialisation formats	I1
P-Rec. 14	Standard vocabularies **SHOULD** be used to describe semantic artefacts	I2
P-Rec. 15	Provenance information regarding the reuse of components from third-party semantic artefacts **SHOULD** be made explicit	I3, R1.2
P-Rec. 16	The semantic artefact **MUST** be clearly licensed for use by machines and humans	R1.1
P-Rec. 17	Provenance **MUST** be clear for both humans and machine	R1.2
Repository		

P-Rec. 4	Semantic Artefact and its content **SHOULD** be published in a trustworthy semantic repository	F4
P-Rec. 5	Semantic repositories **MUST** offer access to Semantic Artefacts and their content using community standard APIs and serialisations to support both use/reuse and indexation by search engines	F4, A1, A1.1
P-Rec. 6	Build semantic artefacts' search engines that operate across different semantic repositories	F4
P-Rec. 7	Repositories **MUST** offer a secure access protocol and appropriate user access control functionalities	A1.2
	Semantic alignment	
P-Rec. 10	Foundational Ontologies **MAY** be used to align semantic artefacts	I1, I2, I3
P-Rec. 11	A standardised knowledge representation language **SHOULD** be used for describing complex logical relations (semantic artefact)	I1
P-Rec. 12	Semantic mappings between the different elements of semantic artefacts **SHOULD** be published in machine-readable formats	I1, I3, R1.3
P-Rec. 13	Crosswalks, mappings and bridging between semantic artefacts **SHOULD** be documented, published, and curated	R1.2, R1.3

1.3.2 *FAIR evaluations*

Since the publication of the FAIR principles, several evaluation methods have been developed to assess FAIRness of data and are listed in Amdouni et al. (Amdouni et al., 2021). To support this assessment, various generic FAIR evaluation tools have been developed such as F-UJI[10], developed in the FAIRsFAIR project, and the FAIRSharing/GOFAIR assessment tool[11] (Wilkinson et al., 2019). These tools have a strong focus on data and may not be suitable for evaluating the FAIRness of ontologies. Therefore, we will consider these tools in a later phase of our work and focus on the existing frameworks for ontologies.

To assess the FAIRness of the ontologies collected in this landscape analysis, we are considering on the one hand the FAIR Semantics recommendations proposed by the FAIRsFAIR project, from which we have defined an evaluation matrix described in section 2.2.2 and on the other hand existing services proposed by other initiatives. By focusing initially on the FAIR Semantics recommendations, we are aiming to contribute practically to the FAIRsFAIR effort and strengthen our collaboration initiated in the context of the Knowledge Exchange Space[12].

Recently, two different FAIRness evaluators for ontologies have been released: O'FAIRe and FOOPS!. O'FAIRe has been proposed by Amdouni et al. (Amdouni et al., 2021) to quantify the level of FAIRness for ontologies stored in any OntoPortal-based repository adopting the MOD (Dutta et al., 2017)

[10] https://www.fairsfair.eu/f-uji-automated-fair-data-assessment-tool

[11] https://fairsharing.github.io/FAIR-Evaluator-FrontEnd/#!/

[12] https://www.ontocommons.eu/news-events/events/creating-knowledge-exchange-space-data-management-and-documentation-kexs-0

metadata model (typically, AgroPortal[13]). The FOOPS! Web service[14] has been published in October 2021 by Garijo et al. (Garijo et al., 2021) and is the first "automatic" FAIRness assessment tool for semantic artefacts. It is based on several relevant works, namely the FAIRsFAIR recommendations and other metadata Best Practices[15] as explained in Poveda-Villalón el al.'s guidelines (Poveda-Villalón et al., 2020).

In this landscape analysis, we have considered two evaluation matrices: the matrix we derived from the FAIR Semantics recommendations and the FOOPs! evaluation matrix. We present in more detail these tools in section 2 and the results of the FAIRness assessment using the two evaluation methodologies in section 3. Other evaluation methodologies and services will be tested in a later phase of our work.

[13] http://agroportal.lirmm.fr/
[14] https://w3id.org/foops/
[15] https://w3id.org/widoco/bestPractices

2. Our approach

We describe in this section the methodology which we have adopted to both collect and analyse domain ontologies for the landscape analysis.

2.1 Surveying the landscape

Our primary objective is to identify as many semantic resources as possible that have been developed in the different domains of interest as defined in section 1.2. We are aware that at the time of the release of this report, the landscape picture will not be complete and that it should be updated during the course of the project. In particular, our task aims at providing a list of semantic resources to populate a dedicated registry which will support the findability of these semantic resources (D3.3).

To start our survey, we leveraged the expertise and knowledge from within the project by collecting ontologies.

This first sampling allowed us to collect 115 ontologies and their domain of coverage. We then organised several events and carried out surveys with domain experts to collect additional ontologies. We describe below each workshop and the related outcomes.

2.1.1 *DORIC-MM: domain ontologies from the pre-workshop survey and MIRO boards*

One of OntoCommons Focused Workshops has contributed towards gathering material for the semantic landscape analysis. This Focused Workshop was named "Domain Ontologies for Research Data Management in Industry Commons of Materials and Manufacturing" (DORIC-MM 2021)" and consisted of two main parts: a preparatory half-day event (kick-off) on the 15th of March, and a full day event (workshop), co-located with the 18th European Semantic Web Conference (ESWC 2021) in June 2021.

The target groups for this activity were both ontologists and Materials and Manufacturing (MM) domain experts. The events were highly interactive, and input was gathered in different ways. Here we mention in particular two sources:

- a short survey embedded with the registration for the March event;
- MIRO boards used during the 15th of March and 7th of June 2021 events.

This allowed us to collect, amongst other things, a list of about 100 domain ontology names and acronyms. This list was cleaned to remove entries that were out of scope or duplicated.

Beside ontology names/acronyms, we also gathered other information, such as examples of current and future use-cases, desiderata/wishes, and various positive experiences in the usage of ontologies.

We have also used a third source, Mentimeter interactive presentations, to collect, for example: opinions on standardisation (in general and in the MM field), opinions and practices in the development of domain ontologies, and a list of relevant projects and initiatives. For further details on all of this, we refer the reader to the dedicated report, OntoCommons first focused workshop on domain ontologies.

When registering for the March event and via subsequent communications, attendees were also invited to fill in a long survey (see the next Section) to describe in detail either the ontologies they develop or an ontology of interest.

2.1.2 *Long ontology survey*

The aim of the long survey was to collect more descriptive and technical metadata about existing ontologies. The outcomes from this survey will not only be used to help outline the domain ontology landscape, but it will also provide the backbone for the documentation "Populated domain ontologies registry". The registry will be created in collaboration with the task "Reference implementation tooling", which will contribute towards the documentation on "OntoCommons ontology registry infrastructure" to make the results of the long survey publicly available.

The long survey is based on experience gained during the READY4SmartCities European project[16], as well as the development of the SmartCities Catalogue (Poveda-Villalón et al., 2014) for ontologies and datasets[17]. It should be noted that OntoCommons will cover ontologies only and not data. The long survey has also taken into account the IoT ontology landscape survey[18] being developed by the AIOTI Semantic Interoperability Expert Group. Finally, the survey was reviewed and complemented with questions related to FAIR principles and their application to ontologies (Hugo et al., 2021; Poveda-Villalón et al., 2020; Garijo et al., 2020; Cox et al., 2021).

This initial survey is available on the OntoCommons website at the following URL: https://ontocommons.eu/node/146. In order to open up the survey to the broadest audience possible, it was accessible without the need to create a dedicated OntoCommons account. For this reason, the form includes an acceptance of the OntoCommons privacy policy and terms of use. This decision was taken as the survey has been designed for everyone (e.g., partners, stakeholders, OntoCommons workshops participants, ontology developers, the broader audience reached by social media, etc.) to be able to enter ontology information at any stage of the project. In particular, a call for filling in the survey by OntoCommons partners has been carried out. The final survey consists of the questions shown in Table 3.

---

[16] https://cordis.europa.eu/project/id/608711

[17] http://smartcity.linkeddata.es/

[18] https://ec.europa.eu/eusurvey/runner/OntologyLandscapeTemplate

Table 3 - Information collected in the ontology long survey

Question	Explanation	Possible values
Name	The name given to the ontology.	Free text
URI	The URI of the ontology.	Free text
Description	A free-text account of the ontology.	Free text
Domains	The different domains covered by the ontology. If the ontology covers more than one domain, please separate them by commas. Example: manufacturing, Materials Science, maintenance, AEC industry, marketing, ...	Free text
Scope	The scope of the ontology in a particular domain e.g., predictive maintenance, stakeholder description, product nomenclature, sensor, building	Free text
Namespace	The preferred namespace URI to use when using terms from this vocabulary.	Free text
Version	The version of the ontology.	Free text
Creation date	The date of formal issuance of the ontology.	Date
Last update	Most recent date on which the ontology was changed, updated, or modified.	Date
Contact person	The person(s) primarily responsible for making the ontology. Please include name and email address of the contact persons whenever possible. If there is more than one contact person, please separate them by commas.	Free text
Publisher	The organisation that published the ontology.	Free text
Ontology language	The ontology language in which the ontology is implemented.	OWL, RDF-S, SKOS, SUO-KIF, Isabelle, (FOL), OBO format, UML, OntoUML, Other
Format	Format in which the ontology code is provided.	RDF/XML, Turtle, N3, N-Triples, TriX, TriG, Other
Use of Top-Level ontologies?	Top level ontologies used by the ontology.	Basic Formal Ontology, DOLCE, SUMO, EMMO, Unified Foundational Ontology, YAMATO, CYC, General Formal Ontology, Other

License	The license of the ontology. Example: CC BY-SA, MIT, etc.	All rights reserved / no license (No Open),CC0 1.0 Universal , CC-BY International, CC-BY Unported, CC-BY-SA International, CC-BY-SA Unported, CC-BY-ND , CC-BY-NC , CC-BY-NC-SA , CC-BY-NC-ND , GFDL, MIT , PDDL, ODC-By, ODBL,W3C software license, Unknown, Other
Please specify (license)	Specify the license if it is not one of the list.	
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).	en – English, es – Spanish, fr – French, de – German, it - Italian, bg – Bulgarian, nl – Dutch, no – Norwegian, ru – Russian, Other
Available documentation	URLs for the documentation of the ontology (for example a website)	Free text
References	Resources that might provide additional information (documents, deliverables, papers, etc.).	Free text
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?	Free text
Best practices	Free text	OBO Foundry, Industry Ontology Foundry principles, FAIR Principles, None, Other
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	Free text

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.	Free text
Is the ontology developed within a standardisation body? If yes, please specify which one	Whether the ontology has been developed in the context of standardisation bodies.	Free text
Is the ontology based on any standards? If yes, please specify which one(s)	Whether the ontology is based on existing standards.	Free text
Is the ontology supported by a community? If yes, please mention the involved community(ies)	Whether the ontology is being supported by any community.	Free text
Is there a sustainability plan for this ontology?	Whether there is a sustainability plan form an organisation, community, company, etc.	None, Yes, No, Maybe, Unknown
Is the ontology being reused by other ontologies or projects? If yes, could please specify which ones?	Whether the ontology is being adopted.	Free text
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.	Free text
Comments	Further information about the ontology that might be relevant.	Free text

At the current point in time, the long survey has received 42 responses out of which 34 are considered as unique and relevant. Answers from the survey have not been extensively studied individually, but we have aggregated it with the dataset, and some of the survey results have been used for the descriptive analysis and the FAIRness evaluation described in sections 3.2 and 3.4, respectively.

2.1.3 *Expert group meeting*

Three group meetings were organised in May 2021 to gather some more ideas about the current state of the domain ontology landscape and to identify gaps. Unlike the DORIC-MM workshop, these meetings were conducted with smaller groups of domain experts and predefined domains were set in the scope of the discussion. The first meeting, held on May 20th, focused on the Physics, Chemistry, and Materials Science domains, whereas two separate meetings were held on May 26th and May 31st, respectively, focusing on Mechanical, Industrial, and Process Engineering. All meetings were held online with a preselected group of invited experts from industry, academia, and research institutes.

Each meeting was held for two hours following a similar format. After a round of introductions, a short presentation was given to participants outlining the agenda and scope of the meeting. The rest of the time was spent in discussion and hands-on activities, for which participants used an online collaborative tool (Miro board) to share their views on the domain landscape.

In the following paragraphs, we first present the agenda of the meetings. We then present the contents of the participants' contributions on the Miro board below, and then provide a synopsis of the discussions.

Agenda:

- Requirement gathering: The goal of this effort was to identify the set of requirements for ontology development in the domain of interest. To reduce the problem further, it is necessary to take the following steps:
  - Identify the focus areas: It is necessary to identify the sub-areas of the domain of interest that needs to be covered by some ontology. Common classification systems, e.g., DFG, ERC may not be suitable from an ontology modelling perspective. Some of the focus areas were already identified from the use cases as they must be covered for the purposes of demonstration by WP5. These areas were communicated to the participants of the meeting for further expansion.
  - Identify connections among focus areas: There are overlaps among the sub-areas. Identification of such overlaps can help us in modularising the ontologies and designing the dependency network (import structure) among different models.
- Identify coverage and gaps based on existing sets of ontologies: In this effort, participants were asked to determine if the focus areas identified in the previous task could be covered by some existing ontology, or if a new ontology needs to be built. Two tasks for this effort were to:
  - Identify one or more ontology from the existing list for each of the focus areas.
  - Identify areas which no existing ontology may cover. Then determine:
    - If an existing ontology could be extended to cover the newly uncovered domain?
    - And if not, should we develop a new ontology for this domain?

Miro board:

The screenshot of the Miro board used in the meeting of May 20th is given in Figure 1. The network diagram is reproduced for better clarity in Figure 2. In Table 4, the list of existing ontologies and their corresponding target domains are presented as identified by the participants.

A single board was used by participants of both meetings, held on May 26th and May 31st respectively. In these two meetings, participants used mind maps to classify the sub-areas. From Figure 3 to Figure 8, each mind map classifies different major areas in sub-areas. In Table 5, the list of existing ontologies and their corresponding target domains are presented as identified by the participants.

A synopsis of the expert workshops' observations are as follows:

- The industry usage of ontologies is still not mature despite their efforts in building ontologies for their own purposes. One of the problems is that engineering thinking is different from ontological thinking.
- Domain ontologies built by academia are varied and sometimes either compete with or complement each other. Some of them are accepted by industry. However, access to these ontologies is greatly hindered by the lack of a library (registry) of these ontologies and associated documentation. On a positive note, these ontologies are usually open when compared to ontologies made by industry as they are not open.
- ISO 15926, which provides an OWL format of their EXPRESS data model for many different facets of manufacturing[19], is quite extensive and exhaustive.
- MatPortal[20] (https://matportal.org/) is a Materials Science ontology portal which seeks to collaborate with OntoCommons efforts.
- A common reference ontology should be built as an upper-level ontology to ensure interoperability; some experts opine that system engineering may be treated as a core ontology for manufacturing in place of a philosophy-driven top-level ontology. Some others think that there is no universal solution for finding one singular model that caters to all areas.
- One single TLO may not be sufficient for ensuring interoperability with every domain level ontology. The efforts of the Industrial Ontology Foundry (IOF) based on BFO are quite promising, but it does not so far necessarily provide an all-in-one complete solution.
- Products and product types are built for procurement also (i.e., for commercial reasons), not only for design. A fully specified product model (from all stakeholder points of view) can therefore be used to create asset models in the design. In design, the components of a product are functional components.
- A process ontology needs to cater for both discrete and continuous manufacturing.

[19] https://www.posccaesar.org/wiki/ISO15926inOWL
[20] https://matportal.org/

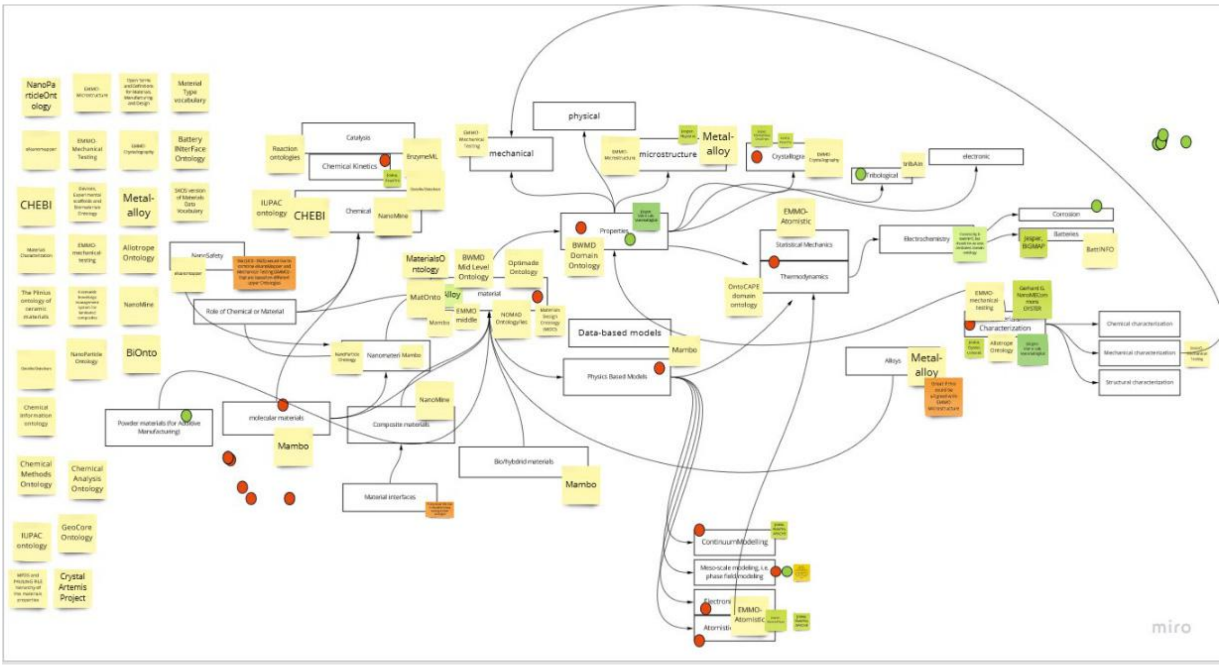

Figure 1 - Frame of a Miro board from the 20th May meeting

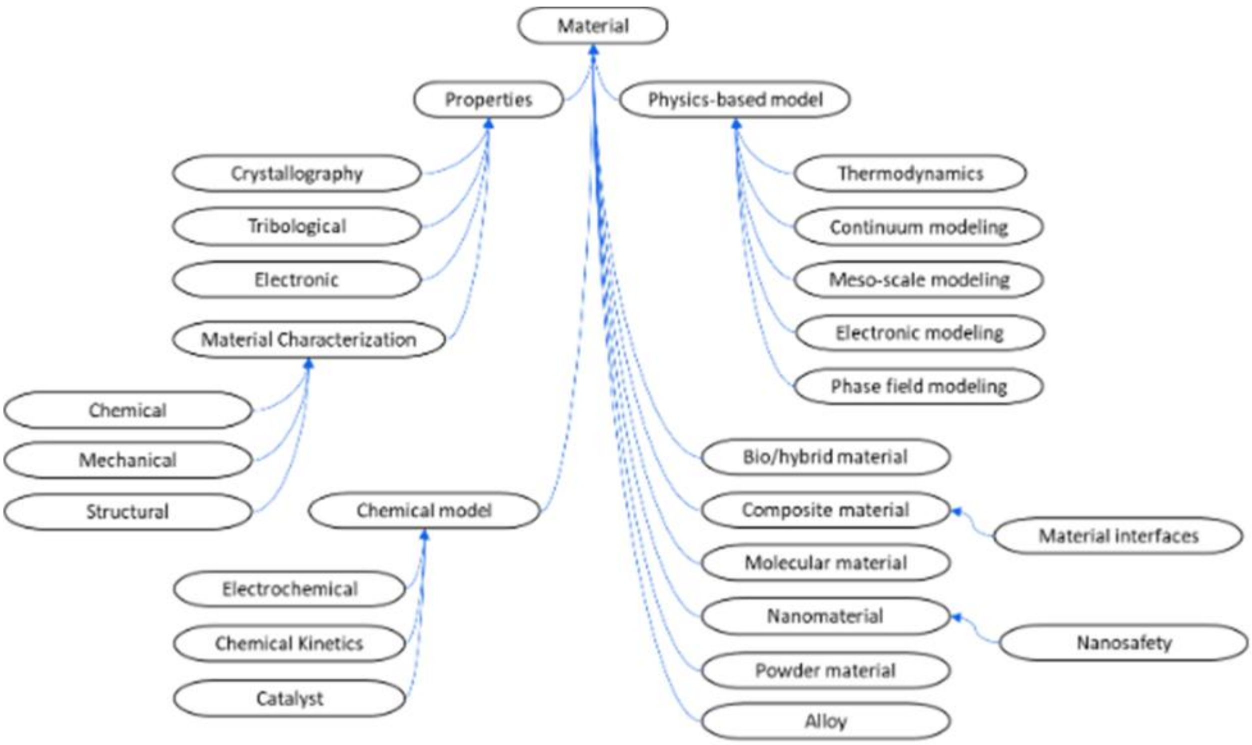

Figure 2 - Focus areas and dependence hierarchy (arrows pointing from area that is required to the area that requires it) (derived from the Miro board in Figure 1)

Table 4 - Landscape mapping for existing ontologies in Physics, Chemistry, and Materials Science

Subject Area	Ontology
Material	BWMD mid-level, Materials ontology, Optimade ontology, Material Design Ontology (MDO), NOMAD, EMMO-middle, MatOnto
Properties	BWMD domain
Mechanical Properties	EMMO-mechanical testing
Microstructure	EMMO-microstructure, PhysMet
Crystallography	Marketplace Ontology, OntoTrans, ReaxPro, EMMO-Crystallography
Tribological	TribAIn
Physics-based Model	Mambo
Thermodynamics	OntoCAPE
ContinuumModeling	ReaxPro, APACHE
Meso-scale modeling	
Atomistic modeling	EMMO-Atomistic, Marketplace, Joana, ReaxPro, APACHE
Electrochemistry	BattInfo
Battery	BattInfo, BIGMAP
Material characterisation	EMMO-mechanical testing, NanoMECommons, OYSTER, Joana, UrWerk, Allotrope ontology, Mat-o-lab, MaterialDigital
Mechanical characterisation	EMMO-mechanical testing
Nanomaterial	NanoParticle ontology, Mambo
Nanosafety	eNanomapper
Composite Material	NanoMine
Material Interface	Many domain ontologies contain some part of this(not particularly identified during the workshop)

Molecular material	Mambo
Bio/hybrid material	Mambo
Chemical model	CHEBI, Ontokin/Ontochem, IUPAC Ontology
Catalyst	Reaction ontology, EnzymeML

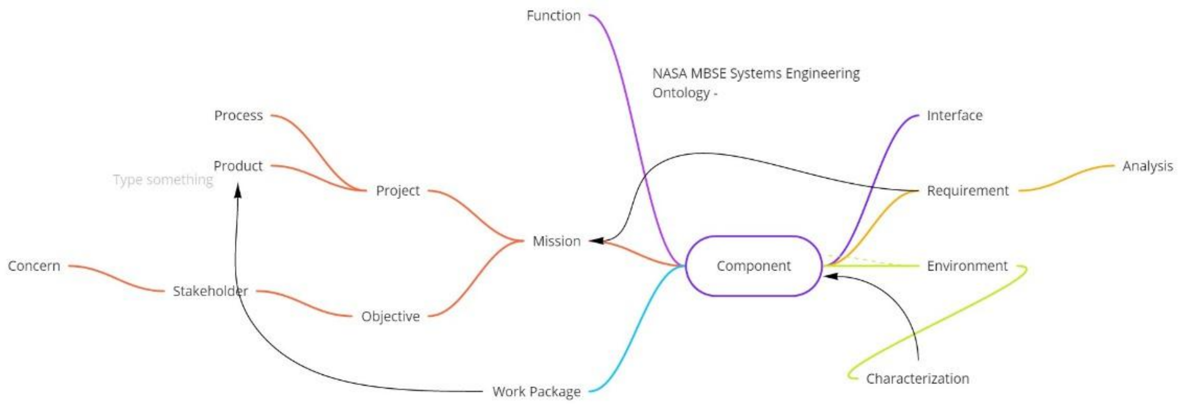

Figure 3 - Mind map for "Component" and sub areas

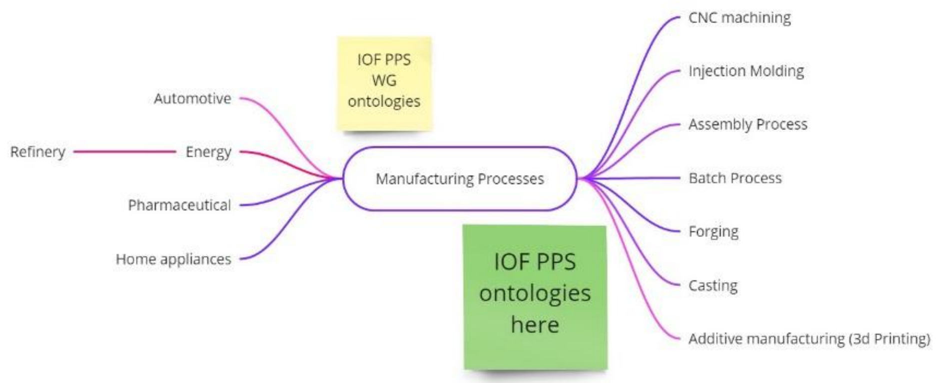

Figure 4 - Mind map for "Manufacturing Process" and its sub areas

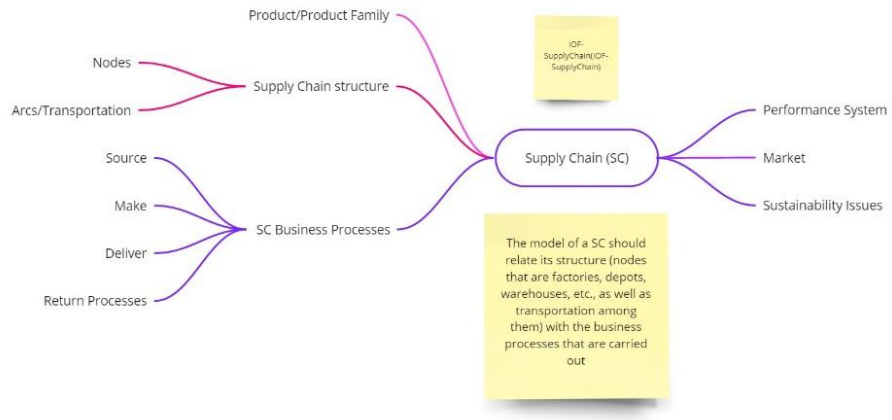

Figure 5 - Mind map for "Supply Chain" and its sub areas

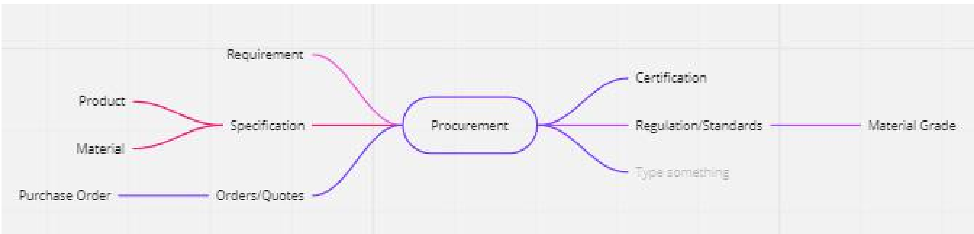

Figure 6 - Mind map for "Procurement" and its sub areas

Figure 7 - Mind map for "Logistic" and its sub areas

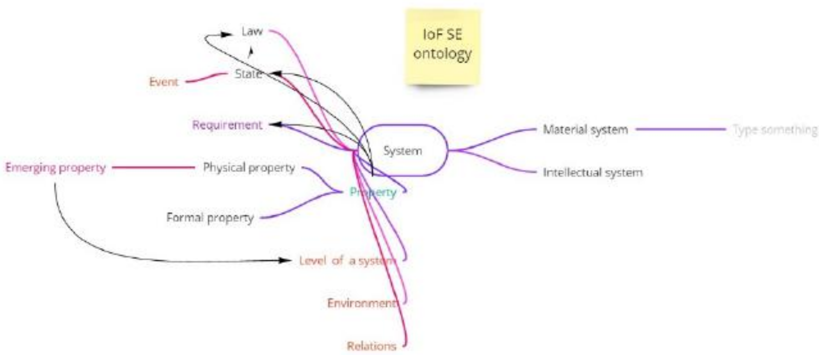

Figure 8 - Mind map for "Logistic" and its sub areas

Table 5 - Landscape mapping for existing ontologies in Mechanical, Industrial, and Process Engineering

Subject Area	Ontology
Design	PRONTO, PSL, SIMPM,
Logistic	IOF-SupplyChain
Manufacturing Process	IOF-PPS
Industrial Maintenance	IOF-Maintenance, ROMAIN
Product	(Product Design), (Product and Process Representation)
Complex System	IOF-PSS, Schedule Reference Ontology (Scheduling Ontology Network)
System	IOF-SE
Supply Chain	IOF-SupplyChain

2.2 Analysing the landscape: our methods

In order to analyse the datasets collected during the initial phase of the landscape survey, we focused on three main aspects:

1. Ontology engineering criteria (format, use of Top-Level Ontologies, type of metadata);
2. FAIRness assessment;
3. Domain criteria (coverage, overlap, semantic gaps, usage, maturity).

We describe in this section the methods used to perform the analysis.

2.2.1 *Ontology engineering criteria*

We consider as ontology engineering criteria various types of information regarding the collected ontologies, namely, their format and serialisation, whether they are machine-readable or non-machine-readable and whether they use Top-Level Ontologies (TLO). In addition, as we will show, we used a topological analysis to examine the inner structure of the ontologies.

2.2.1.1 **High level analysis**

To perform this analysis, we have manually inspected the ontologies available in machine readable format and collected information regarding the TLO used, the format, the serialisation and we compiled this information into an Excel spreadsheet. Results of this high-level analysis have been completed and validated by some of the results of the topological analysis (format, syntax, … ; see below). Results are presented in section 3.

2.2.1.2 Topological analysis: methods and tools

Machine-readable domain ontologies collected from various workshops and surveys were subjected to automatic measurement tools to extract a set of topological metrics. Topological metrics of an ontology include basic measurements, such as number of axioms, classes, properties, and individuals but also complex measurements such as average number of subclass and superclass axioms, and if the ontology contains cycles[21]. Qualitative measurements include those such as the level of expressivity and its subscriptions to different types of OWL and RDFS profiles, such as DL, RL, and QL. The list of features that are measured as part of this analysis is given in Table 6. Topological metrics provide a view on the shape of the ontology in terms of not only its breadth and depth but also how expressive the model is and in which format it is encoded. Several topological measures on individual concepts are used in calculating the similarity and ultimately produce alignment among them. However, metrics are presented as aggregates for the whole model.

The metrics are calculated by reusing a tool called ROBOT (Jackson et al., 2019), which was developed to analyse and edit Open Biomedical Ontologies. ROBOT is an open-source Java library and command-line tool. The 'measure' command computes some or all the available metrics for this tool based on the given argument for 'mode of calculation' (Essential, Extended, All, Reasoner). However, ROBOT does not provide bulk processing. A separate script has therefore been written in Java to compute the metrics for a set of ontologies in bulk. This task was performed in three phases: pre-processing, computation, and analysis. Table 6 contains a few terms and abbreviations, taken from the ROBOT library, that might need further explanation:

- **ABOX**: The part of the ontology that is about individuals
- **RBOX:** The part of the ontology that is about object properties
- **TBOX**: The part of the ontology that is about classes
- **AVG**: average
- **MAX:** maximum
- **RHS**: ride-hand-side. If used in conjunction with axioms, usually the "superclass" part of a subClassOf axiom.
- **DT**: datatype
- **GCI**: General concept inclusion. Formally, all subClassOf axioms are GCIs, but here we mean those that have a complex left-hand-side (sub-class part).
- OWL 2 Profiles:
    - **OWL 2 DL**: Roughly the subset of OWL that conforms to description logics, plus annotations. It is geared towards ontologies with a high degree of expressivity. Reasoning tasks can be relatively expensive.
    - **OWL 2 EL**: Subset of OWL 2 DL, which is geared towards scalable reasoning in the TBox.

[21] https://www.w3.org/TR/owl-ref/#equivalentClass-def

- **OWL 2 QL**: Subset of OWL 2 DL, which is geared towards scalable query answering in the ABox.
- **OWL 2 RL**: Subset of OWL 2 DL that can be handled using logic programs.

- **EXPRESSIVITY**: A sequence of letters that characterise the logical features used by the ontology. A breakdown of the meaning of the letters is described in section 3.2.

In the pre-processing phase, all ontologies that have a machine-readable format available at a permanent URL are shortlisted from the main collection. It should be noted that despite some being described in peer-reviewed articles, a machine-readable source for all the identified ontologies could not be acquired at the time of writing this report (see Section 3). The shortlisted ontologies are then stored in a comma-separated value file (CSV) with three columns: # (serial number), Name (a short title for the ontology), and Source (the URL of the machine-readable format).

In the computation phase, the CSV file with short-listed ontology sources is read into the Java script. The script uses the ROBOT library to retrieve the measurement. The result is then written into a result file. Some of the ontologies could not be loaded by the library for various reasons (mostly because there was a failure in loading the imported ontology). These ontologies and the associated errors generated during loading were stored in a separate log file.

In the analysis phase, a Python script is used to read the results CSV generated from the Java script to generate different plots as shown in section 3. In this analysis, a few additional measures were calculated from others. These are listed in Table 7.

Table 6 - Topological metric features and their descriptions

Feature	Description
AXIOM_COUNT	Number of axioms
AXIOM_COUNT_INCL	Number of axioms (including all imported ontologies)
INDIVIDUAL_COUNT	Number of individuals
ANNOTATION_PROPERTY_COUNT	Number of annotation properties
ANNOTATION_PROP_COUNT_INCL	Number of annotation properties (including all imported ontologies)
DATATYPE_NOTBUILTIN_COUNT	Total number of distinct custom (not built-in) datatypes.
DATATYPE_NOTBUILTIN_COUNT_INCL	Total number of distinct custom (not built-in) datatypes (including all imported ontologies)
DATATYPE_BUILTIN_COUNT	Total number of distinct built-in datatypes
OBJPROPERTY_COUNT	Number of object properties
ONTOLOGY	
OBJPROPERTY_COUNT_INCL	Number of object properties (including all imported ontologies)
DATATYPE_COUNT	Total number of distinct datatypes

DATATYPE_COUNT_INCL	Total number of distinct datatypes (including all imported ontologies)
ONTOLOGY_ID	ontology IRI
MAX_NUM_NAMED_SUPERCLASS	Maximum number of named superclasses
MAX_NUM_NAMED_SUPERCLASS_INCL	Maximum number of named superclasses (including all imported ontologies)
CYCLE	If true, there is a definite cycle in the ontology. If false, cyclicity is unknown
CYCLE_INCL	If true, there is a definite cycle in the ontology. If false, cyclicity is unknown (including all imported ontologies)
MAX_AXIOMLENGTH	Longest axiom in terms of number of entities used
MAX_AXIOMLENGTH_INCL	Longest axiom in terms of number of entities used (including duplicate uses) (including all imported ontologies)
CLASS_SGL_SUBCLASS_COUNT	Number of super-classes which have more than one subclass
CLASS_SGL_SUBCLASS_COUNT_INCL	Number of super-classes which have more than one subclass (including all imported ontologies)
INDIVIDUAL_COUNT_INCL	Number of individuals. (including all imported ontologies)
TAUTOLOGYCOUNT	Number of tautological axioms.
TAUTOLOGYCOUNT_INCL	Number of tautological axioms (including all imported ontologies)
AVG_ASSERT_N_SUPERCLASS_INCL	Average number of (asserted) superclasses per class (including all imported ontologies)
AXIOM_COMPLEXRHS_COUNT	Number of axioms with a complex right-hand side
AXIOM_COMPLEXRHS_COUNT_INCL	Number of axioms with a complex right-hand side (including all imported ontologies)
BOOL_PROFILE_OWL2	Does the ontology correspond to the OWL2 profile?
BOOL_PROFILE_OWL2_DL	Does the ontology correspond to the OWL2 DL profile?
SIGNATURE_SIZE	Total number of entities in signature, including classes and individuals.
SIGNATURE_SIZE_INCL	Total number of entities in signature, including classes and individuals (including all imported ontologies)
MULTI_INHERITANCE_COUNT_INCL	Number of classes with multiple inheritance. (including all imported ontologies)
GCI_COUNT_INCL	Total count of GCI

GCI_HIDDEN_COUNT	Total count of hidden GCIs.
GCI_HIDDEN_COUNT_INCL	Total count of hidden GCIs. (including all imported ontologies)
SYNTAX	The serialisation that is used for the ontology
AVG_ASSERT_N_SUBCLASS_INCL	Average number of (asserted) subclasses per class (including all imported ontologies)
ABOX_SIZE	Number of axioms in the ABox
ABOX_SIZE_INCL	Number of axioms in the ABox (including all imported ontologies)
LOGICAL_AXIOM_COUNT	Number of logical axioms in ontology
LOGICAL_AXIOM_COUNT_INCL	Number of logical axioms in ontology (including all imported ontologies)
EXPRESSIVITY	Logical expressivity
EXPRESSIVITY_INCL	Logical expressivity (including all imported ontologies)
TBOXRBOX_SIZE	Number of TBOX axioms (with RBOX)
TBOXRBOX_SIZE_INCL	Number of TBOX axioms (with RBOX) (including all imported ontologies)
CLASS_COUNT	Number of classes
CLASS_COUNT_INCL	Number of classes (including all imported ontologies)
AVG_INSTANCE_PER_CLASS	Average number of individuals per class
AVG_INSTANCE_PER_CLASS_INCL	Average number of individuals per class (including all imported ontologies)
AVG_ASSERT_N_SUPERCLASS	Average number of (asserted) superclasses per class
AVG_ASSERT_N_SUBCLASS	Average number of (asserted) subclasses per class
BOOL_PROFILE_OWL2_EL	Does the ontology correspond to the OWL2 EL profile?
BOOL_PROFILE_RDFS	Does the ontology correspond to the RDFS profile?
BOOL_PROFILE_OWL2_RL	Does the ontology correspond to the RL profile?
BOOL_PROFILE_OWL2_QL	Does the ontology correspond to the QL profile?
MOST_FRQUENTLY_USED_CONCEPT	The most frequently used class
MULTI_INHERITANCE_COUNT	Number of classes with multiple inheritance
RBOX_SIZE	Number of axioms in the RBOX
RBOX_SIZE_INCL	Number of axioms in the RBOX (including all imported ontologies)
TBOX_SIZE	Number of axioms in the TBOX

TBOX_SIZE_INCL	Number of axioms in the TBOX (including all imported ontologies)
DATAPROPERTY_COUNT	Total number of distinct data properties
DATAPROPERTY_COUNT_INCL	Total number of distinct data properties (including all imported ontologies)
RULE_CT	Number of SWRL rules used.
UNDECLARED_ENTITY_COUNT	Number of undeclared entities
UNDECLARED_ENTITY_COUNT_INCL	Number of undeclared entities (including all imported ontologies)
TBOX_CONTAINS_NOMINALS	Number of TBOX axioms with nominals
TBOX_CONTAINS_NOMINALS_INCL	Number of TBOX axioms with nominals (including all imported ontologies)

Table 7 - Additional features used in the analysis that are calculated from the features in Table 6

Feature	Expression	Description
ANNOTATION_PROP_PER_CLASS	ANNOTATION_PROP_COUNT / CLASS_COUNT	Number of annotation properties per class. This gives an indication about how well the classes are documented.
TBOX_SIZE_PER_CLASS	TBOX_SIZE / CLASS_COUNT	The relative size of the TBOX per class. This give a picture of how many class-related axioms there are per class in the ontology.
ABOX_SIZE_PER_CLASS	ABOX_SIZE / CLASS_COUNT	The relative size of the ABOX per class. This

2.2.2 *FAIRness assessment: method and tools*

As no existing framework for evaluating the FAIRness of semantic artefacts exists based on the FAIRsFAIR recommendations, we have analysed the "FAIR Semantics" recommendations and have summarised into the following table whether the recommendation is mandatory (red), recommended (yellow), optional (green) or unspecified (grey), as well as whether they apply to ontologies, semantic artefact repositories or whether they pertain to the semantic community at large.

Table 8 - FAIR Semantics recommendations by priority level and target

Rec #	Recommendation	Target
P-Rec 1	Globally Unique, Persistent and Resolvable Identifiers MUST be used for Semantic Artefacts, their content (terms/concepts/classes and relations), and their version	Ontology
P-Rec 2	Globally Unique, Persistent and Resolvable Identifiers MUST be used for Semantic Artefacts metadata records. Metadata and data must be published separately, even if it is managed jointly	Ontology/Repository
P-Rec 3	A common minimum metadata schema MUST be used to describe semantic artefacts and their content	Ontology
P-Rec 5	Semantic repositories MUST offer access to Semantic Artefacts and their content using community standard APIs and serialisations to support both use/reuse and indexation by search engine	Repository
P-Rec 7	Repositories MUST offer a secure access protocol, and appropriate user access control functionalities	Repository
P-Rec 8	Human and machine-readable persistence policies for semantic artefacts metadata and data MUST be published	Repository
P-Rec 9	Semantic artefacts MUST be made available as a minimum portfolio of common serialisation formats	Ontology/Repository

P-Rec	**16**	The Semantic Artefact MUST be clearly licenced for use by machines and humans	Ontology
P-Rec	**17**	Provenance MUST be clear for both humans and machines	Ontology
P-Rec	**4**	Semantic artefacts and its content SHOULD be published in a trustworthy semantic repository	Ontology
P-Rec	**11**	A standardised knowledge representation language SHOULD be used for describing semantic artefacts	Ontology
P-Rec	**12**	Semantic mappings between the different elements of semantic artefacts SHOULD be published in machine readable format	Semantic Community
P-Rec	**13**	Crosswalks, mappings and bridging between semantic artefacts SHOULD be documented, published, and curated	Semantic Community
P-Rec	**14**	Standard vocabularies SHOULD be used to describe semantic artefacts	Ontology
P-Rec	**15**	Provenance information regarding the reuse of components from third-party semantic artefacts SHOULD be made explicit	Ontology
P-Rec	**10**	Foundational Ontologies MAY be used to align semantic artefacts	Ontology
P-Rec	**6**	Build semantic artefact search engines that operate across different semantic repositories	

To evaluate the degree of FAIRness of ontologies, we chose to define three levels: 1) Fully FAIR, which would fulfil all the recommendations, even the optional ones; 2) Minimally FAIR, which would fulfil only the mandatory recommendations; and 3) Not FAIR, expressed as percentage of compliance with the mandatory recommendations. In total, we have selected 9 relevant recommendations (4 mandatory, 4 recommended and 1 optional).

For this work, we consider only the recommendations which are directly related to the Semantic Artefacts (SA):

- Usage of Globally Unique, Persistent and Resolvable Identifiers for a semantic artefact and its content (P-Rec 1);
- Usage of Globally Unique, Persistent and Resolvable Identifiers for a version (P-Rec 1);
- Usage of descriptive Metadata (P-Rec 3);
- Vocabularies used for describing semantic artefact (P- Rec 14);
- Ontology Publication (P-Rec 4);
- Licence (P-Rec 16);
- Usage of provenance (P-Rec 17 and P-Rec 15);
- Usage of TLO (P-Rec 10);
- Ontology Knowledge Representation Language (P- Rec 11).

For each of these aspects, we have defined dedicated questions to be considered for the evaluation. The questions are listed in the table below. These questions should have a simple yes/no answer and a FAIR Semantic Artefact should have a yes answer to all the questions.

Table 9 - FAIR Semantics evaluation questions

Rec #	Topic	Question
P-Rec 1	GUPRI	Does the SA have a persistent identifier of type purl, w3id or handle except for DOI?
P-Rec 1	GUPRI	Does the identifier resolve to a machine-readable format?
P-Rec 1	GUPRI	Does the SA provide a GUPRI for version?
P-Rec 3	Metadata	Does the SA have descriptive metadata?
P-Rec 14	Standard Vocabularies	Does the SA's metadata use widely used vocabularies (dc, dct, etc.)?
P-Rec 17	Provenance	Does the SA have provenance information?
P-Rec 17	Provenance	Does the SA use W3C PROV?
P-Rec 15	Provenance	Does the SA describe imports with provenance?

P-Rec 4	Publication	Is the SA published on a dedicated trusted semantic repository?
P-Rec 16	Licence	Does the SA have a license?
P-Rec 16	Licence	Is the license machine-readable?
P-Rec 11	Language	Does the SA use a standard knowledge representation such as SKOS, OWL, etc.?
P-Rec 10	TLO	Does the SA align with a Top-Level Ontology?

In total, we have 13 questions that could be answered with yes/no. To be considered as compliant to a recommendation with multiple related questions, the Semantic Artefact must have a yes answer to all the related questions. If it does not, we will consider in this first version of the analysis that the recommendation is not fulfilled.

The analysis performed with this particular grid has been done manually, and has leveraged some information collected by the topological analysis. (See the section above for more information and section 3 for the results). This work was tedious and error prone and cannot be extended to the whole dataset. We therefore considered existing services and their associated evaluation methodology.

As we mentioned in the introduction, two other evaluation approaches have been proposed to evaluate the FAIRness of ontologies: O'FAIRe and FOOPS!. These two approaches have been implemented as services to automate the evaluation.

O'FAIRe provides a list of questions for each FAIR principle based on existing evaluation frameworks for data such as the RDA FAIR Data Maturity Model (FDMM, Bahim et al., 2020), the RDA "SHAring Reward and Credit" FAIRness assessment grid (SHARC, David et al., 2020), and existing recommendations for semantic artefacts such as the FAIRsFAIR recommendations, and Garijo et al.'s guidelines and best practices (Garijo et al., 2020), MIRO (Matentzoglu et al., 2018). These questions are associated with a score and a list of FAIR metadata that can be used to determine how much of an ontology is FAIR or FAIRer. The questions can be found on GitHub in both text[22] and JSON-LD[23] format. This evaluation methodology has been implemented within a service associated with AgroPortal which could not be used at the time of the writing for supporting our analysis. The number of questions considered in the O'FAIRe evaluation matrix is rather high (62 questions). To apply it for our analysis would require tedious manual work and many calculations. Therefore, we have chosen to evaluate this methodology in a later stage of our work.

FOOPS! Is the first "automatic" FAIRness assessment tool for semantic artefacts. It is based on several relevant works, namely the FAIRsFAIR recommendations and other metadata Best Practices[24] as explained in Poveda el al.'s guidelines (Poveda-Villalón et al., 2020). The web service takes an OWL

[22] https://github.com/agroportal/fairness/blob/master/doc/results/FAIR-questions.md
[23] https://github.com/agroportal/fairness/blob/master/src/main/resources/config/common/questions.config.json
[24] https://w3id.org/widoco/bestPractices

ontology (or SKOS thesauri) as an input and generates a total FAIR score that is calculated from the 24 different checks covering all of the FAIR principles. The 24 tests are carried out as follows: 9 checks for Findable, 3 checks for Accessible, 3 checks for Interoperable and 9 checks for reusable. Each test has a number, evaluates a specific aspect, and could be answered by a yes/no. All the tests are described in Table 10 below.

Table 10 - FOOPS! tests

Test #	Topic	Description
PURL1	Identifier	Persistent URL. This check verifies if the ontology has a persistent URL (W3ID, PURL, DOI, or a W3C URL).
URI1	Identifier	Ontology URI resolvable. This check verifies if the ontology URI found within the ontology document is resolvable.
VER1	Identifier	Version IRI. This check verifies if there is an id for this ontology version, and whether the id is unique (i.e., different from the ontology URI).
VER2	Identifier	Version IRI resolves. This check verifies if the version IRI resolves.
URI2	Identifier	Consistent ontology IDs. This check verifies if the ontology URI is equal to the ontology ID.
OM1	Metadata	Minimum metadata. This check verifies if the following minimum metadata [title, description, license, version IRI, creator, creationDate, namespace URI] are present in the ontology.
FIND1	Metadata	Ontology prefix. This check verifies if an ontology prefix is available.
FIND2	Registry	Prefix is in the registry. This check verifies if the ontology prefix can be found in prefix.cc or LOV registries. This check also verifies if the prefix resolves to the same namespace prefix found in the ontology.
FIND3	Registry	Ontology in metadata registry. This check verifies if the ontology can be found in a public registry (LOV).
CN1	Identifier	Content negotiation for RDF and HTML. This check verifies the ontology URI is published following the right content negotiation for RDF and HTML.
FIND_3-BIS	Metadata	Metadata is accessible, even when ontology is not. Metadata are accessible even when the ontology is no longer available. Since the metadata is usually included in the ontology, this check verifies whether the ontology is registered in a public metadata registry (LOV).

HTTP1	Protocol	Open protocol. This check verifies if the ontology uses an open protocol (HTTP or HTTPS).
RDF1	Ontology	RDF availability. This checks if the ontology has an RDF serialisation (ttl, n3, rdf/xml, json-ld).
VOC1	Metadata	Vocabulary reuse (metadata). This check verifies if the ontology reuses other vocabularies for declaring metadata terms.
VOC2	Metadata	Vocabulary reuse. This check verifies if the ontology imports/extends other vocabularies (besides RDF, OWL and RDFS).
DOC1	Provenance	HTML availability. This check verifies if the ontology has an HTML documentation.
OM2	Metadata	Recommended metadata. This check verifies if the following recommended metadata [NS Prefix, version info, creation date, citation] are present in the ontology. It also checks if “contributor” is present, but with no penalty (as not all ontologies may have a contributor).
OM3	Metadata	Detailed metadata. This check verifies if the following detailed metadata [DOI, publisher, logo, status, source, issued date] are present in the ontology. It also checks if other metadata [previous version, backward compatibility, modified] are present, but with no penalty (as not all ontologies may have, e.g., a previous version).
VOC3	Metadata	Documentation labels. This check verifies the extent to which all ontology terms have labels (rdfs:label in OWL vocabularies, skos:prefLabel in SKOS vocabularies).
VOC4	Metadata	Documentation definitions. This check verifies whether all ontology terms have descriptions (rdfs:comment in OWL vocabularies, skos:definition in SKOS vocabularies).
OM4.1	License	License availability. This check verifies if a license is associated with the ontology.
OM4.2	License	License is resolvable. This check verifies if the ontology license is resolvable.
OM5.1	Provenance	Basic provenance metadata. This check verifies if basic provenance [author, creation date] is available for the ontology. This check also verifies whether further metadata [contributor, previous version] are present, but with no penalty (as not all ontologies may have a previous version or a contributor).

OM5.2	Provenance	Detailed provenance metadata. This check verifies if detailed provenance information is available for the ontology: issued date, publisher.

Based on the results of the tests, FOOPS! generates a final FAIR score that indicates how much a semantic artefact complies with the FAIR principles for example, a score of 100% means that an ontology is respecting all of the FAIR principles. As an example, Figure 9 illustrates an overview of the SAREF extension for industry and manufacturing domain ontology results.

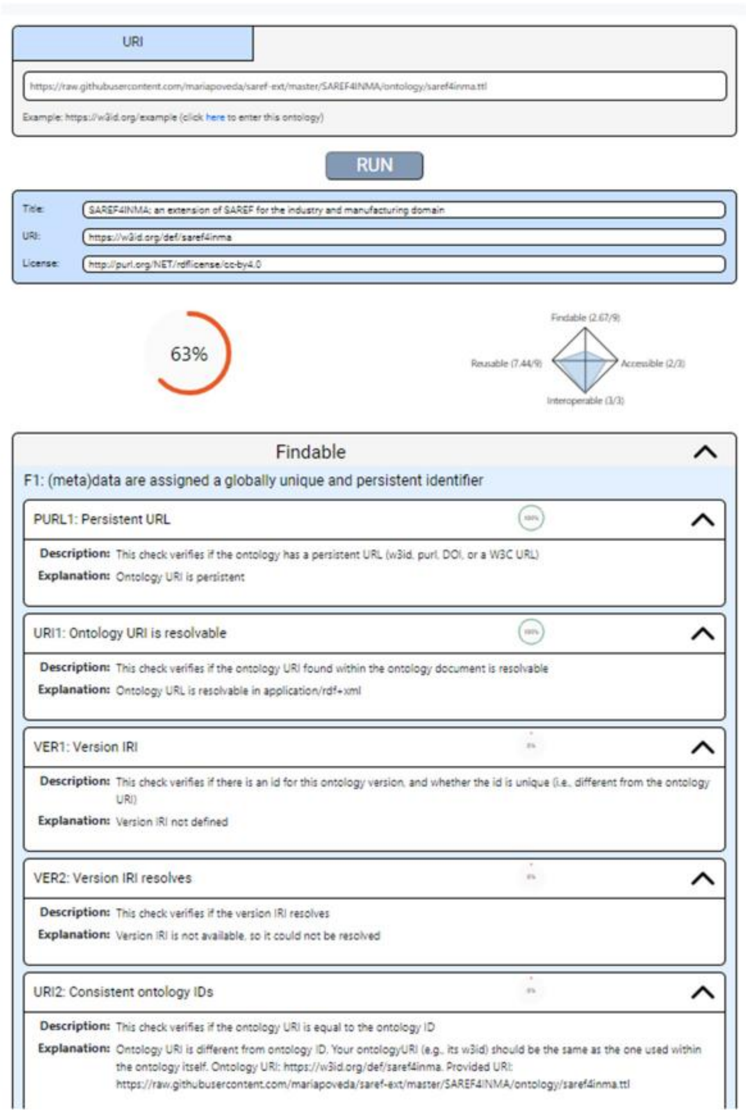

Figure 9 - Analysis results for SAREF using FOOPS!

For our analysis, we will use both our evaluation methodology based on the FAIRsFAIR recommendations and the FOOPS! web service. Our objective is to compare the scores obtained by both approaches and evaluate their divergence. Ultimately, our objective is to work towards the convergence of these methodologies.

2.2.3 *Automating our analysis and making ontologies more FAIR*

One of the main challenges we faced with our analysis and, in particular, with the domain coverage, ontology overlap, and FAIRness assessment was with the programmatic access to the semantic artefact content. To support our approach, we are leveraging an ontology repository developed by NCBO and branded as the OntoPortal virtual appliance 25. OntoPortal is a well-known fully Semantic Web compliant technology offering several key features (e.g., ontology mapping, annotator, recommender, etc.) and it is supported by a consortium of researchers in the OntoPortal Alliance. The NCBO technology is domain-independent and open-source. It is implemented for several domains: biomedicine (BioPortal26 located in the USA), agronomy (AgroPortal27 located in France), environment (EcoPortal28 located in Italy). In addition, the virtual appliance enables anyone to deploy an ontology server and customise it based on specific needs. OntoPortal provides a rich API to access concepts and relations within the published ontologies. On the one hand, uploading and publishing ontologies in such repository increases the FAIRness of the semantic artefacts by complying to the FAIR Semantics recommendations (specifically P-Rec 4). On the other hand, it provides us with a vital service to access programmatically the ontologies identified in our landscape survey so as to perform complex analysis and FAIRness assessment.

2.2.3.1 **MatPortal: an ontology repository for Materials Science**

During a session of the RDA Vocabulary and Semantic Service Interest Group in 2021, we discovered MatPortal[29], an OntoPortal instance to store and publish Materials Science ontologies. We immediately established contact. This resource is currently developed and maintained by Fraunhofer in Germany and is becoming an important part of several projects including the German National Research Infrastructure, NFDI. As the content of this new portal overlaps strongly with our area of interest, we have established a collaboration framework in which we will provide the relevant ontologies we identified in the landscape analysis to enrich the content of MatPortal.

2.2.3.2 **IndustryPortal: a common ontology portal for industry and related domains**

As mentioned earlier, several semantic artefacts (i.e., vocabularies, taxonomies, and ontologies) exist to represent and annotate data in the industry and its related domains. The more ontologies are created in the domain, the more the need to have a common infrastructure to facilitate identification, reuse, alignment, and maintenance become necessary. Therefore, we propose a common ontology portal for industry to deal with all of the knowledge management issues and serve the community in the long term. One of the main expected outcomes of the OntoCommons project is to develop a reference ontology repository for the industry domain.

[25] https://ontoportal.org/the-ontoportal-virtual-appliance/
[26] https://bioportal.bioontology.org/
[27] http://agroportal.lirmm.fr/
[28] http://ecoportal.lifewatch.eu/
[29] https://matportal.org/

In that context, we have developed and deployed the IndustryPortal[30], a prototype version of an ontology repository located in France at ENIT (École Nationale d'Ingénieurs de Tarbes). Currently, IndustryPortal hosts 30 ontologies which are not hosted in other NCBO-based portals; we are working regularly to import new ones by contacting the original users. Figure 10 shows a screenshot from the summary page of MASON (MAnufacturing's Semantics ONtology) in the IndustryPortal.

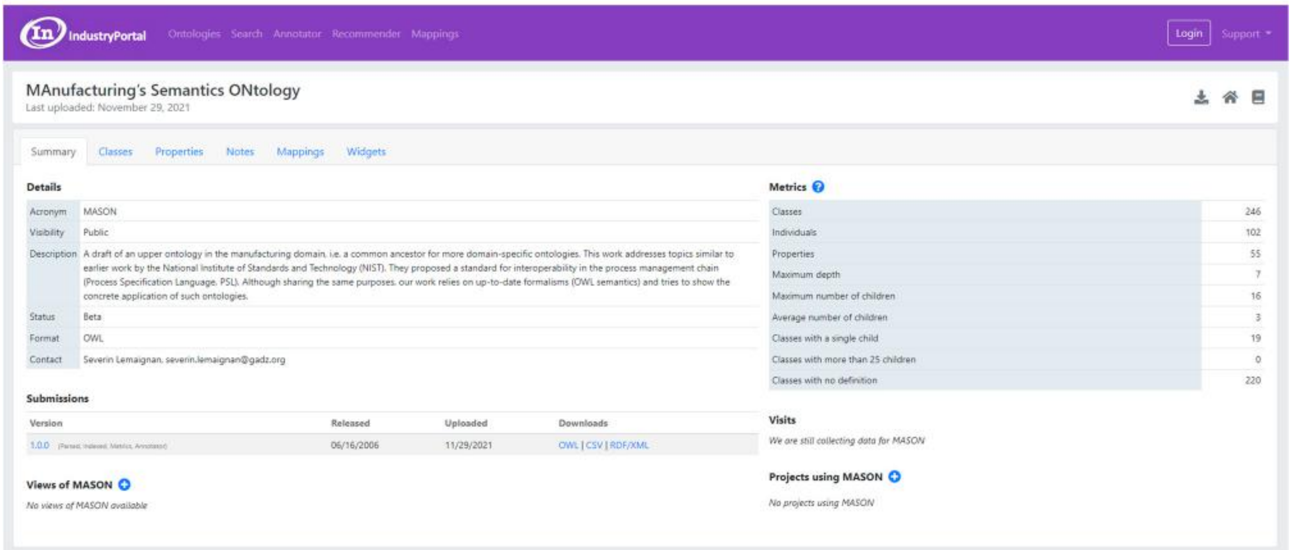

Figure 10 - Industry Portal User Interface

The current version of the prototype enables users:

- To submit a new ontology in public or private mode;
- To edit various ontology metadata (i.e., acronym, visibility, description, status, format, and contact);
- To save all versions of an ontology;
- To search and browse across all the ontologies;
- To annotate a piece of text with all the ontologies;
- To store and serve mappings between ontologies (inside and outside the portal).

In the future, we will load more ontologies, enrich our ontology metadata description model, enlarge the number of portal users, and enable exchange with other existing OntoPortal initiatives (mainly MatPortal, BioPortal and AgroPortal). In addition, we will extend the functionalities and user interfaces by adding, for example, a FAIRness assessment module based on the code of the O'FAIRe Web service[31] proposed by AgroPortal. This integration will enable an automatic computation of all the FAIR scores for hosted ontologies in the IndustryPortal along with a comprehensive graphic visualisation (e.g., wheel/pie charts, histograms, etc.) of the obtained scores.

[30] http://industryportal.enit.fr/
[31] https://github.com/agroportal/fairness

3. Results

We summarise in this section the results of our analysis.

3.1 Dataset

In the initial gathering phase we collected 222 entries out of which we identified initially a total of 130 ontologies (88 machine-readable ontologies, including 9 from MatPortal, and 42 non-machine-readable ontologies). This dataset will evolve and be enriched by additional ontologies during the course of the project.

The ontologies have been collated into an Excel spreadsheet and annotated with additional information such as the type of TLO used or the associated domain, the serialisation format, etc.

3.2 Domain distribution of the dataset

The first criterion we examined was the distribution of ontologies by domains for the whole dataset. This distribution is shown below. The number of ontologies is comprised of ontologies that were found to be described in scientific articles and ontologies that were available on the Web. These ontologies are classified according the initial categories defined previously (see section 1.2). When classifying ontologies within our dataset, we identified ontologies that are related to industry but which would not fit within the initial categories due to their more general scope (e.g., related to management, product customer services, etc.). We chose to keep them in the dataset and to create an additional category named "Other".

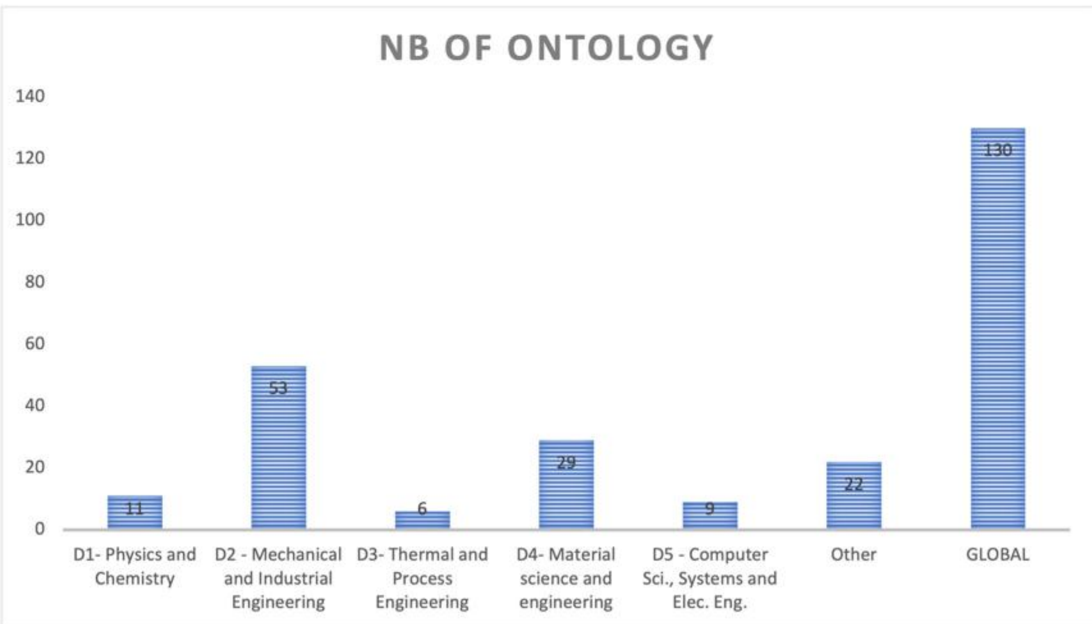

Figure 11 - Distribution of ontologies by domains

This representation shows that the most prominent domain in terms of ontologies is the Mechanical and Industrial Engineering domain (53 ontologies), followed by Materials Science and Engineering (29 ontologies). For the other domains, it is most likely that fewer resources are available. However, we cannot guarantee that we have not under-sampled these domains.

3.3 Ontology engineering criteria

In this section we investigate in more details the collected ontologies (i.e., format, serialisation, machine readable vs. non machine readable, …) for each domain.

3.3.1 *Not machine-readable vs machine actionable*

During the collection phase, we realised that some of the ontologies listed in D3.1 and provided by experts either do not exist or are only described in paper format, e.g., in scientific publications. However, to be FAIR and reusable, ontologies must be published in a machine-readable format along with their documentation (e.g., reports, papers, etc.). We therefore investigated the proportion of machine-readable ontologies by manually checking the associated URL (either collected in the long survey or during workshops) and contacted authors of the papers to get (wherever possible) machine-readable versions. These machine-readable ontologies have been added directly to our list. Out of the 130 relevant ontologies, 88 of them are machine readable, which is a good score in the field.

It is interesting to notice that the domain of Thermal and Process Engineering does not have any machine-readable ontologies at the time of the writing. This is a gap that should be tackled in the future.

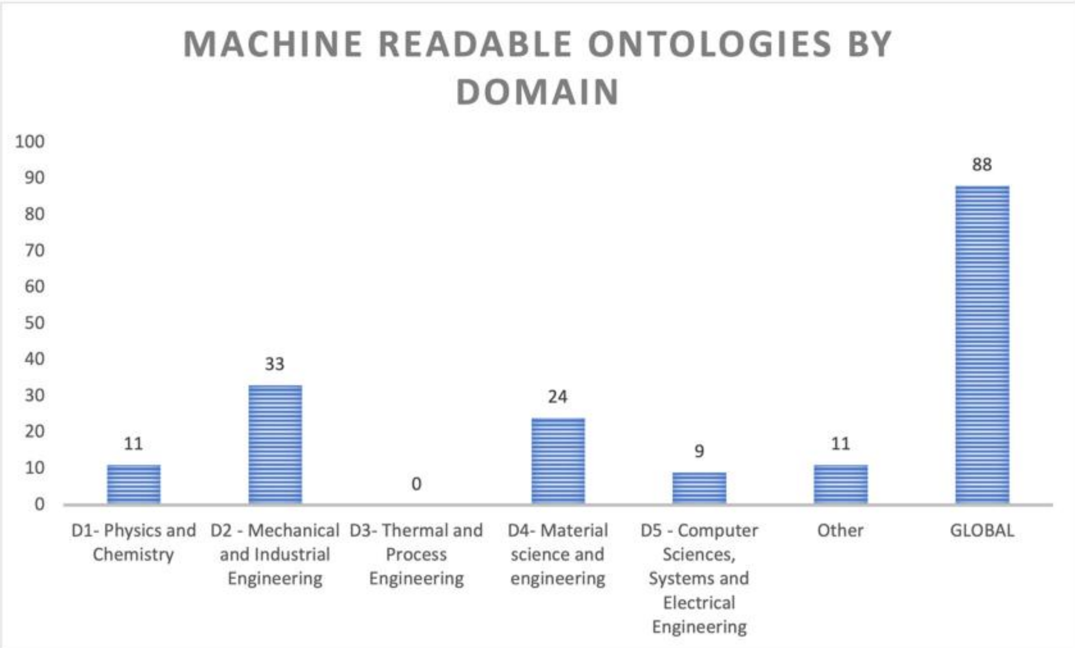

Figure 12 - Distribution of Ontologies per domain

We investigated the ratio of machine-readable versus non machine-readable ontologies by domains. The result of this analysis are shown below.

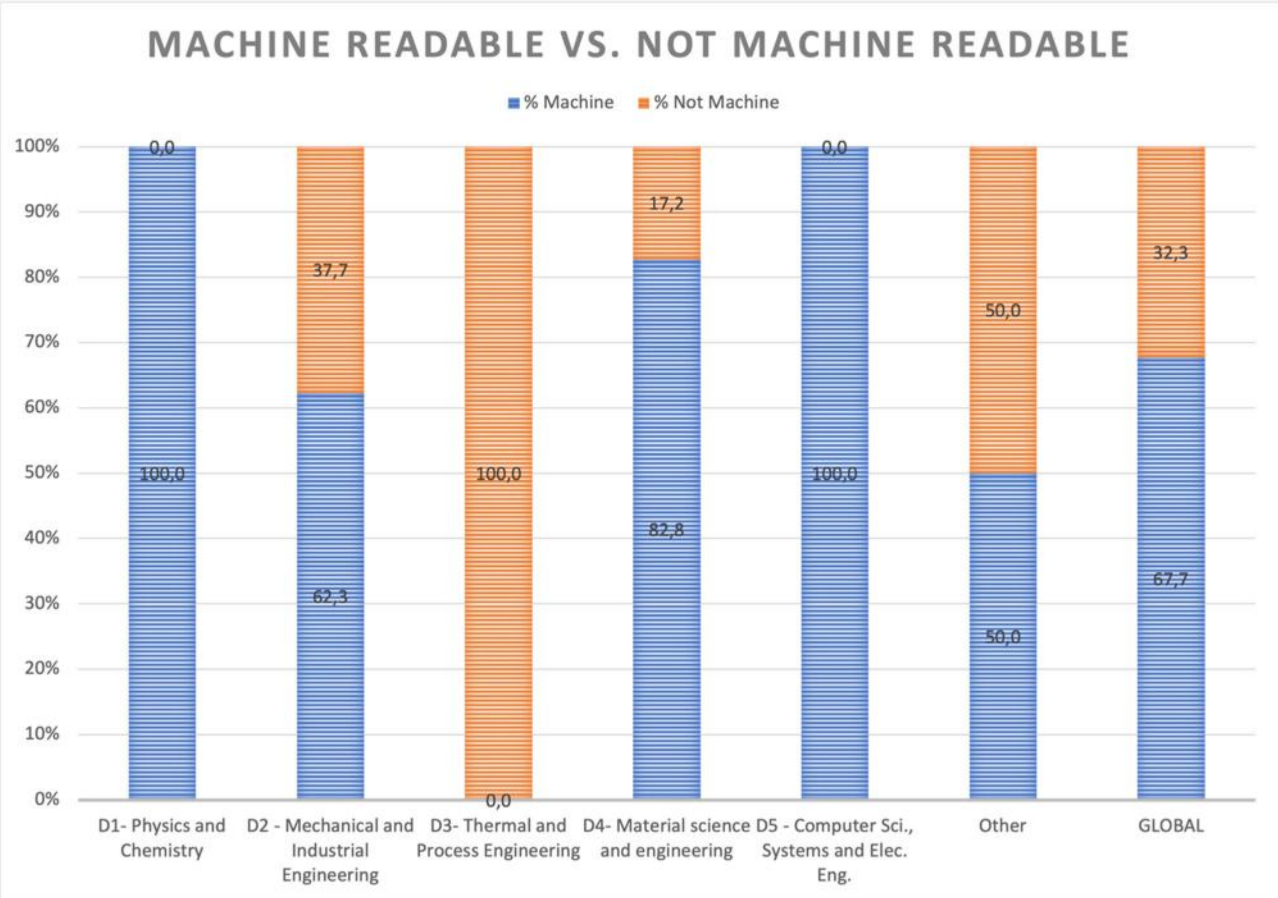

Figure 13 - Machine readable format vs not machine-readable format

Except for the domain of “Physics and Chemistry” for which all the ontologies we collected are in a machine-readable form, all domains have a portion of non-machine-readable ontologies. This represents 1/3 of the dataset in the global. As mentioned, at the current point in time, the domain of “Thermal and Process Engineering” has no machine-readable ontologies identified.

3.3.2 *Alignment with Top Level Ontologies*

In this analysis, we investigated the usage of different Top-Level Ontologies (TLOs) within the ontology pool. For this, we manually analysed the ontology file to define whether the ontologies were leveraging a TLO.

We discovered that among the corpus of ontologies considered, two main TLO were used: the Basic Formal Ontology (BFO) and the Elementary Multiperspective Material Ontology (EMMO). A third ontology, the Smart Applications REFerence (SAREF), has been added to this analysis. However, it is important to note that SAREF is not a TLO but rather a mid-level ontology.

We show in Figure 14 below the percentage of ontologies in each domain aligned with a TLO and the percentage including all domains (global). In the overall list of ontologies, without any domain consideration, we found that 33% of the ontologies are aligned with TLOs. The Physics and Chemistry domain has a maximum percentage (73%) of ontologies aligned to a TLO.

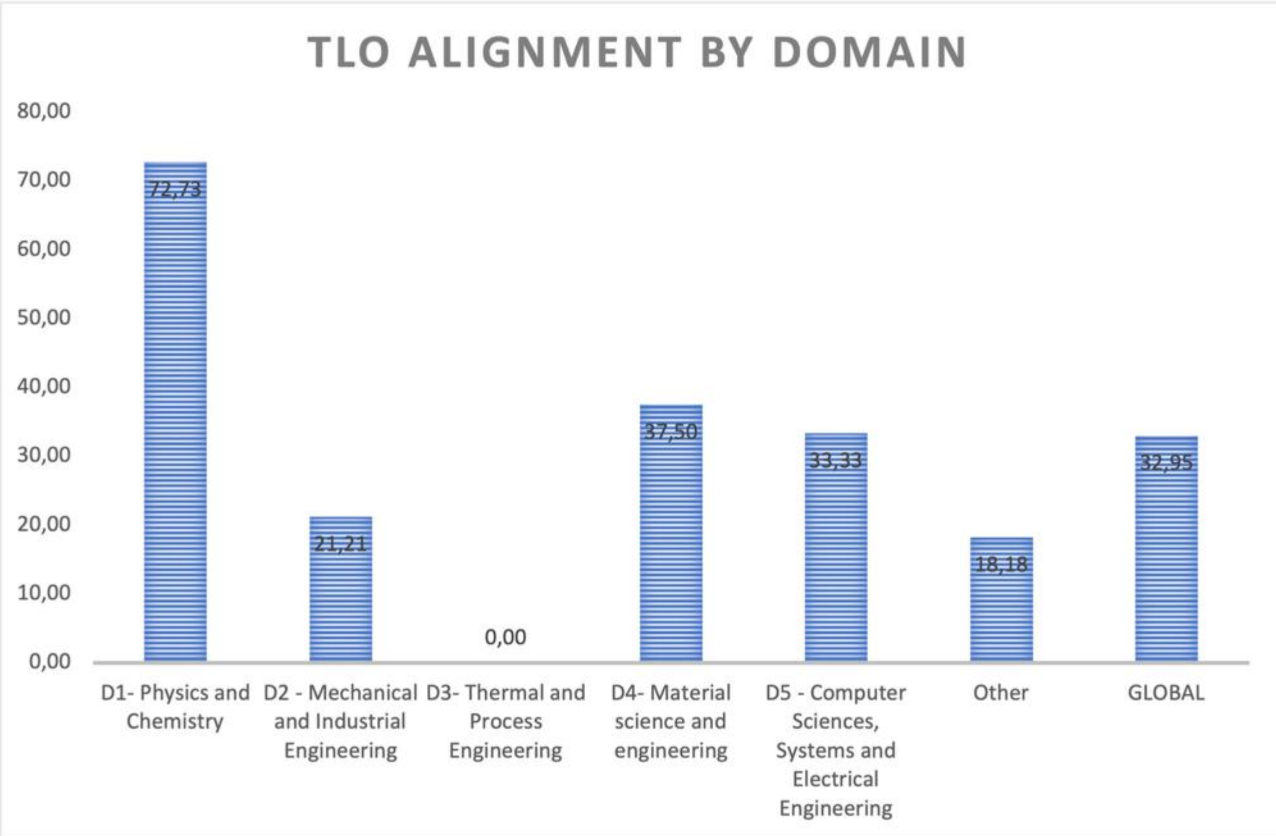

Figure 14 - Alignment with Top Level Ontologies

In Figure 15 below, we show the distribution of TLO usage. It reveals that the most widely used TLO in our dataset is BFO (23%), followed by EMMO (7%).

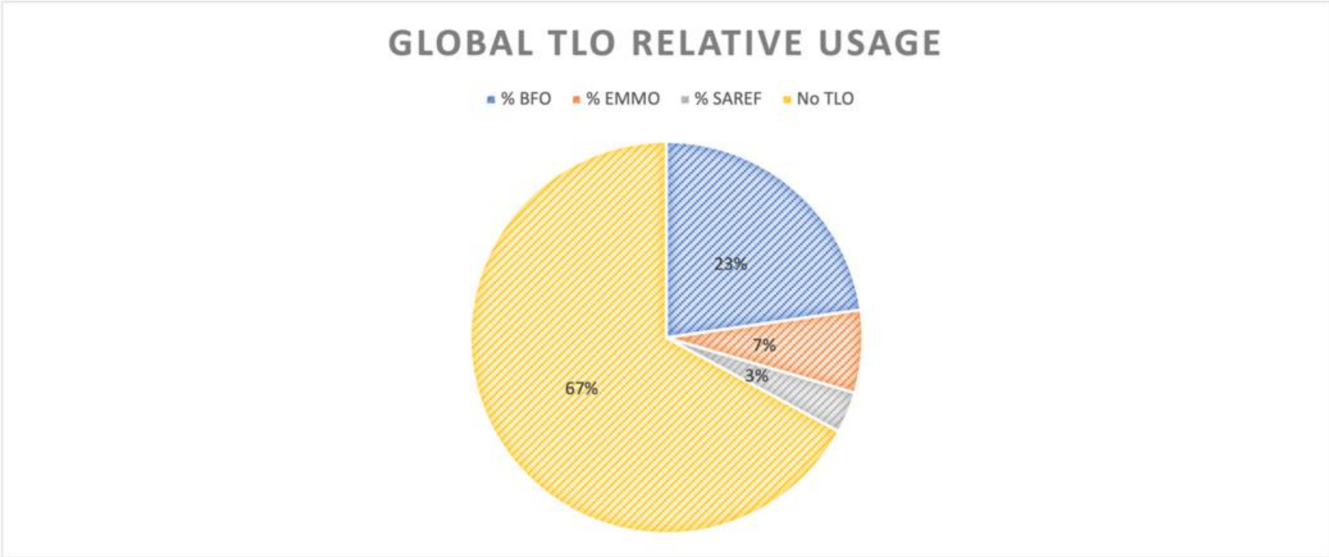

Figure 15 - Distribution of TLO usage for the whole dataset

In Figure 16 below, we show the distribution of TLO usage in the four domains in which we have machine readable ontologies. We are not considering in this figure the ontologies categorised as "Other" because these ontologies are more generic.

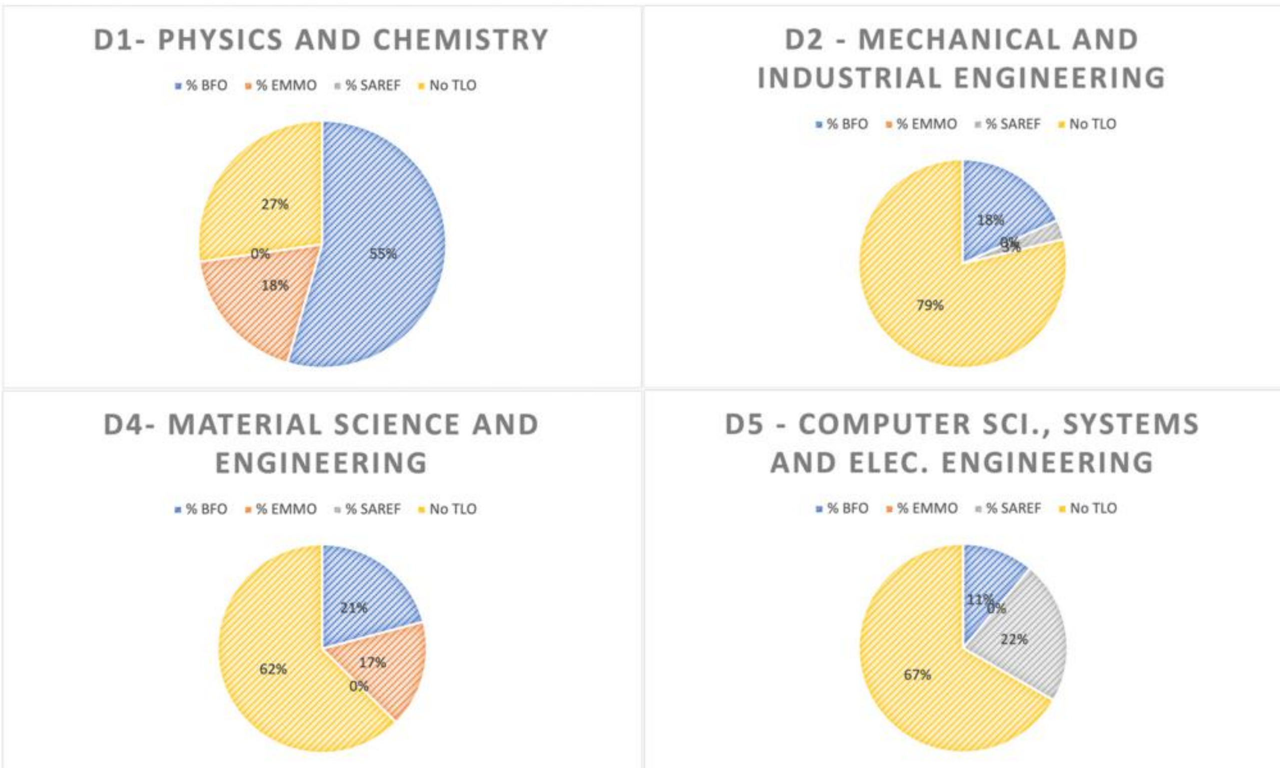

Figure 16 - Distribution of TLO usage by domain

3.3.3 *Serialisation format*

In this analysis, we investigate the various serialisation formats used to publish ontologies. We identified three main serialisations: RDF/XML, OWL/XML and Turtle. In some case, multiple serialisations were proposed for the same artefact. We then considered a fourth category called MultiSyntax. The results are shown below.

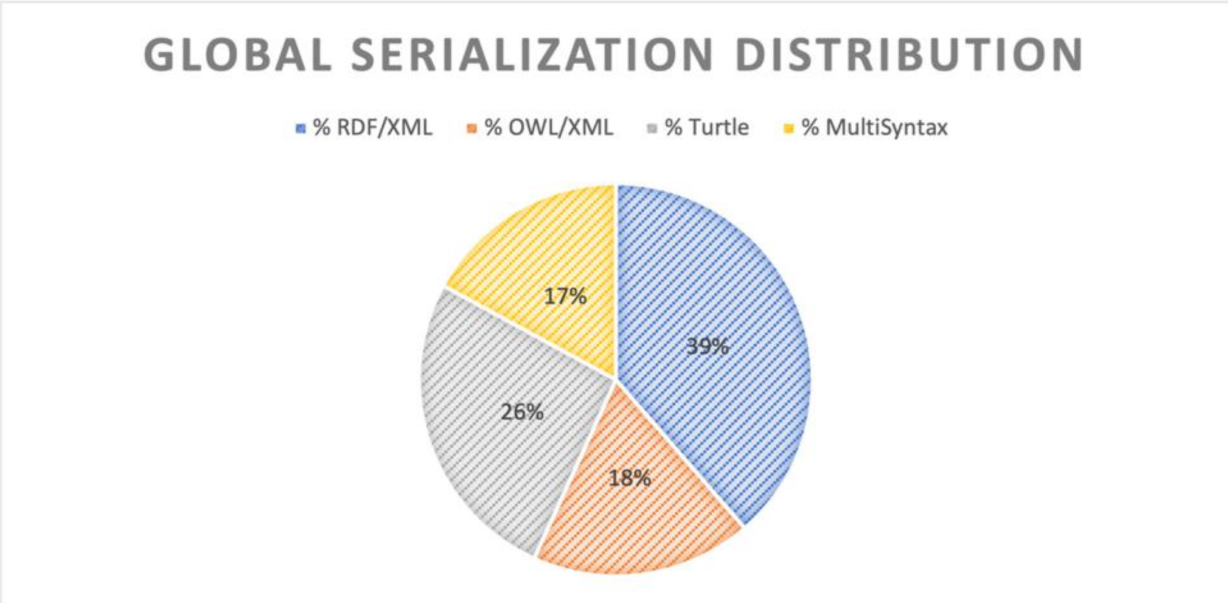

Figure 17 - Distribution of the serialisation format for the whole dataset (global)

At a global level, we see that the main format used to publish machine-readable ontologies is RDF/XML followed by Turtle and OWL/XML. This distribution is not preserved across the different domains considered, i.e., in D1 – Physics and Chemistry and D5- Computer Science, Systems and Electrical Engineering for which there are no ontologies using OWL/XML syntax. D4 - Materials Science and Engineering is the domain for which we find a majority of semantic resources offered with various serialisation formats (MultiSyntax).

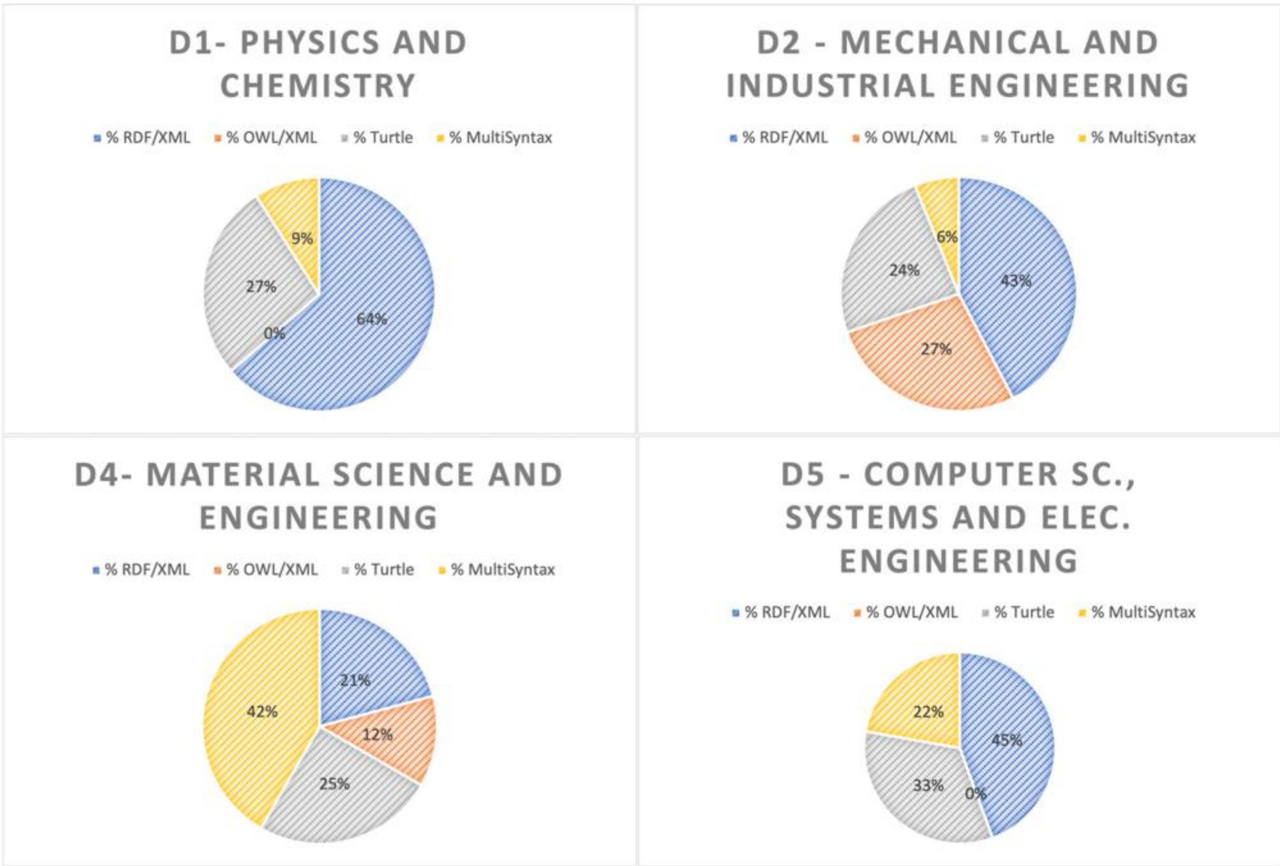

Figure 18 - Distribution of serialisation format by domain

3.3.4 *Topological Analysis*

3.3.4.1 **Domain**

Based on the discussions and feedback received in the workshops, the domain ontologies were grouped into five domains, as summarised in Figure 19. As mentioned previously, several ontologies such as ScorVoc and Semantic Types Ontology did not match this categorisation, and were categorised as *other*. The colours associated with each category are reused in the analysis in section 3.

- Physics & Chemistry
- Material Science and Engineering
- Computer science, systems and electrical engineering
- Mechanical and industrial engineering
- Other

Figure 19 - Identified domains and their associated colours used in the following plots.

3.3.4.2 Numerical metrics

A set of numerical metrics for each domain ontology included in this analysis as plots in Figure 20 to Figure 26 (see Table 15 and Table 16 in Section 6 for complete numerical data).

In Figure 20, we see that the size of the ontologies (in terms of the number of axioms) varies a lot from ontology to ontology. The largest is CHEBI, covering over 2.5 million axioms. The Universal Standard Product and Services Classification (USPSC) is another large ontology. However, unlike CHEBI which mainly defines classes, USPSC defines a large number of individuals, which might be unexpected for a domain ontology.

While looking at other aspects of ontology size in Figure 21 and Figure 22, we see that the Coordinated Holistic Alignment of Manufacturing Processes, MatOnto and Allotrope ontologies have a significantly higher annotation count compared to the others. When it comes to object properties and data properties, Schema.org is the dominant ontology. For data type count, it is harder to get a clear picture as the variation is relatively small. For RBox-size, the Allotrope, Coordinated Holistic Alignment of Manufacturing Processes and ScorVoc ontology have a significantly larger size in comparison to the others.

If we look at the relative measures, shown in Figure 23 and Figure 24, the picture changes. The Organization and Sensor Observation Sampling Actuator ontologies both have more than one annotation property per class. The LinkedDesign ontology has by far the largest TBox size per class, while the Standards ontology is dominating the ABox size per class, and the average instance per class.

Figure 25 and Figure 26 show measures of the complexity of different ontologies. We see that the average number of asserted sub- and superclasses is close to one for all ontologies in all domains. Just a few ontologies have a slightly higher average number of sub- and superclasses, including CHEBI and EMMO Mechanical Testing. The average number of superclasses with more than one subclass per class is about 10% overall and also for each of the domains. However, within the domains there are some variations. In particular, EMMO Atomistic stands out with an average number of superclasses having more than one subclass of 50%. Multi-inheritance is used in 35% of the analysed ontologies and most can be found in the NanoParticleOntology. Multi-inheritance is most frequently used in the Physics and Chemistry and Materials Science and Engineering domains. However, some ontologies within Mechanical and Industrial Engineering use multi-inheritance rather intensively (in terms of multi-inheritance per class) in ontologies like Product Life Cycle Engine and Scheduling Reference Ontology. The overall average number of axioms with complex right-hand side per class is 52%. A few ontologies, like ExtruOnt uses it more extensively, but there are also several ontologies with no complex right-hand side.

In Figure 27 to Figure 29 we try to compare the different domains by comparing the average value of different metrics for different domains in a spider plot. From Figure 30 we can see that all domains are using OWL2 DL, but in the computer science domain, OWL2 RL is also popular.

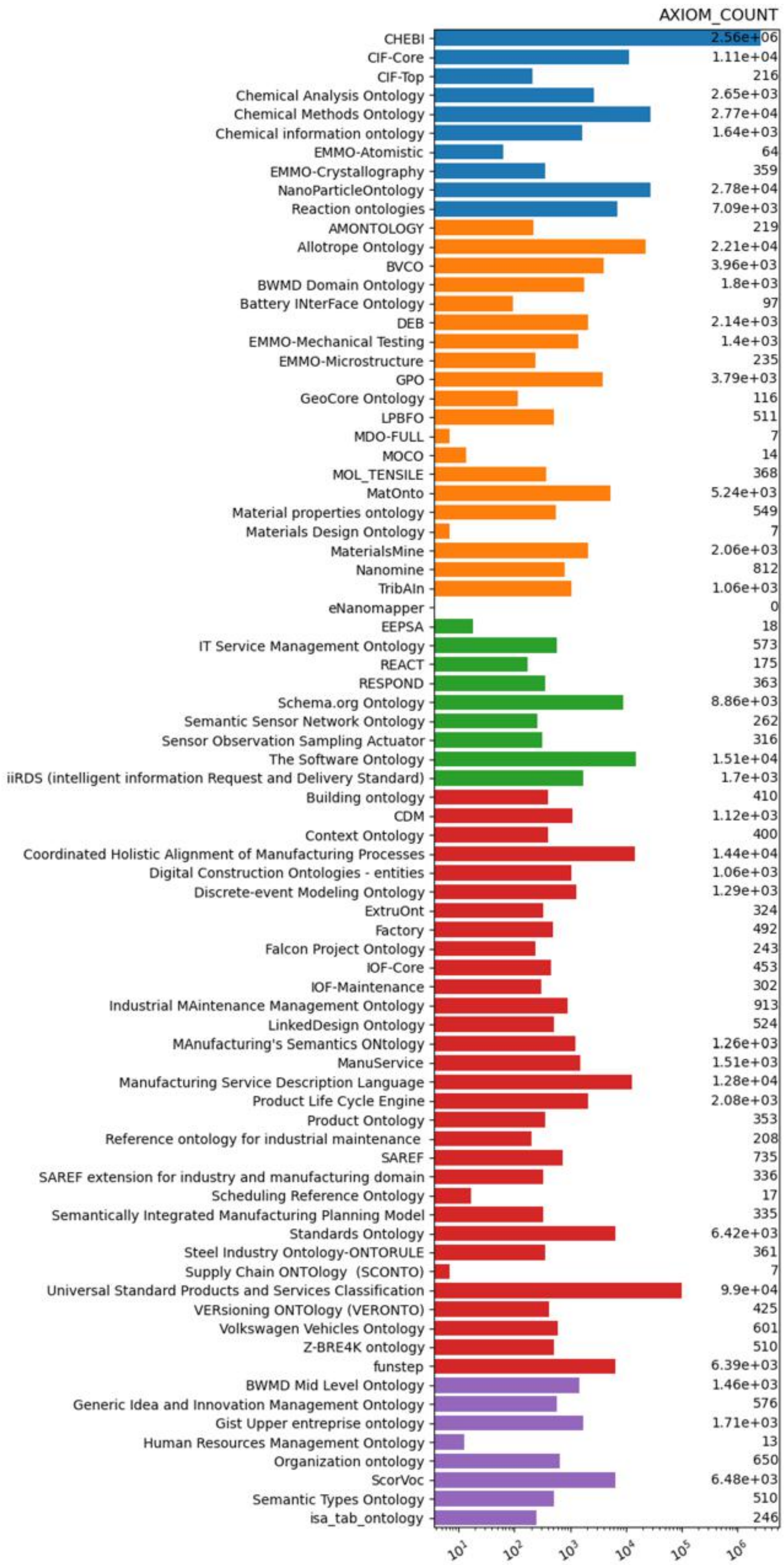

Figure 20 - Size of the analysed ontologies in terms of the total number of axioms they contain. The colours correspond to domains according to Figure 19. Note the logarithmic scale.

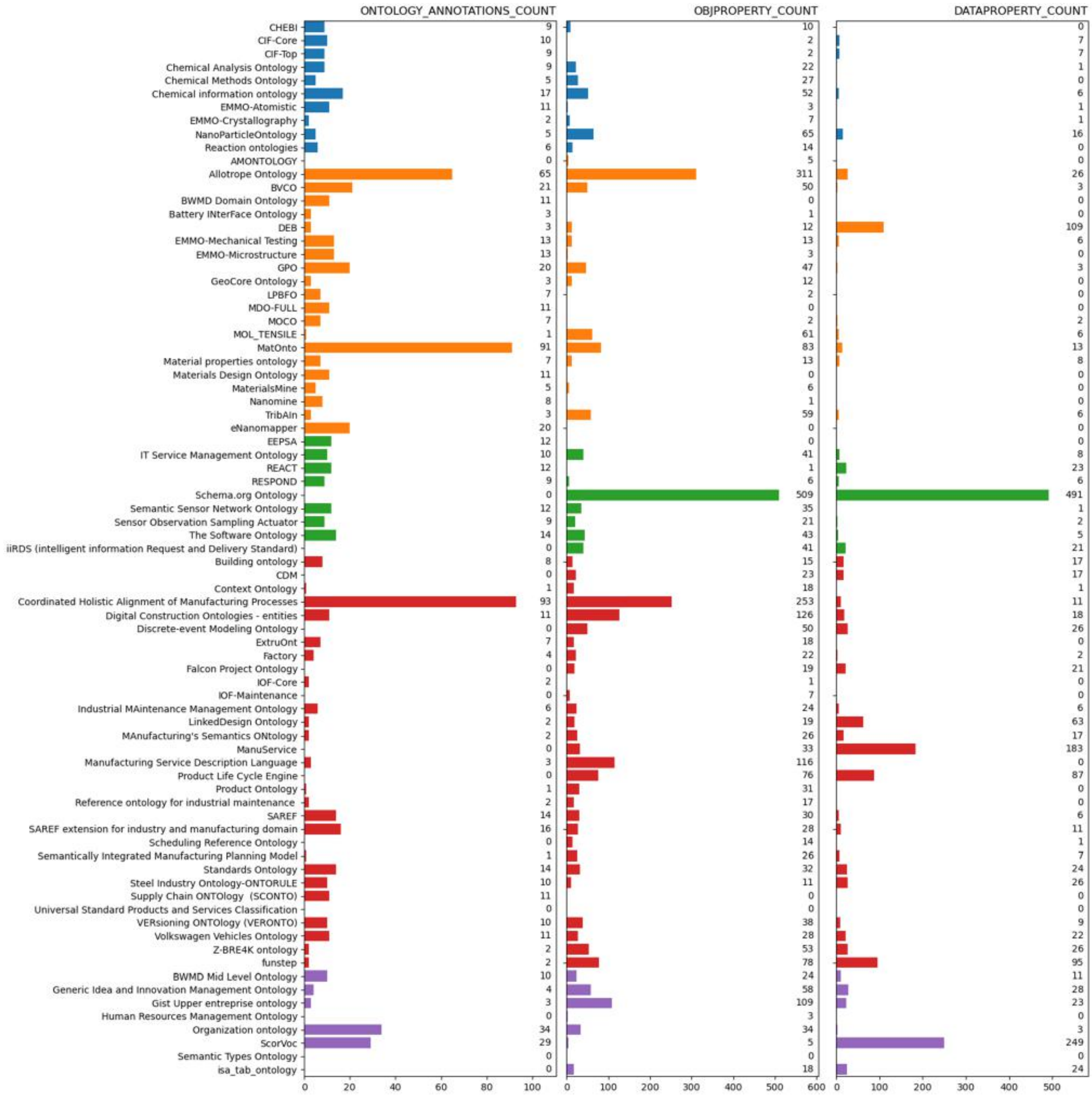

Figure 21 - Other measures of the size of the ontologies in terms of the number of annotations, object properties and data properties. The colours correspond to domains according to Figure 19.

Figure 22 - Other measures of the size of the ontologies in terms of the number of data types and axioms in the RBox (the part of the ontology that is about object properties). The colours correspond to domains according to Figure 19.

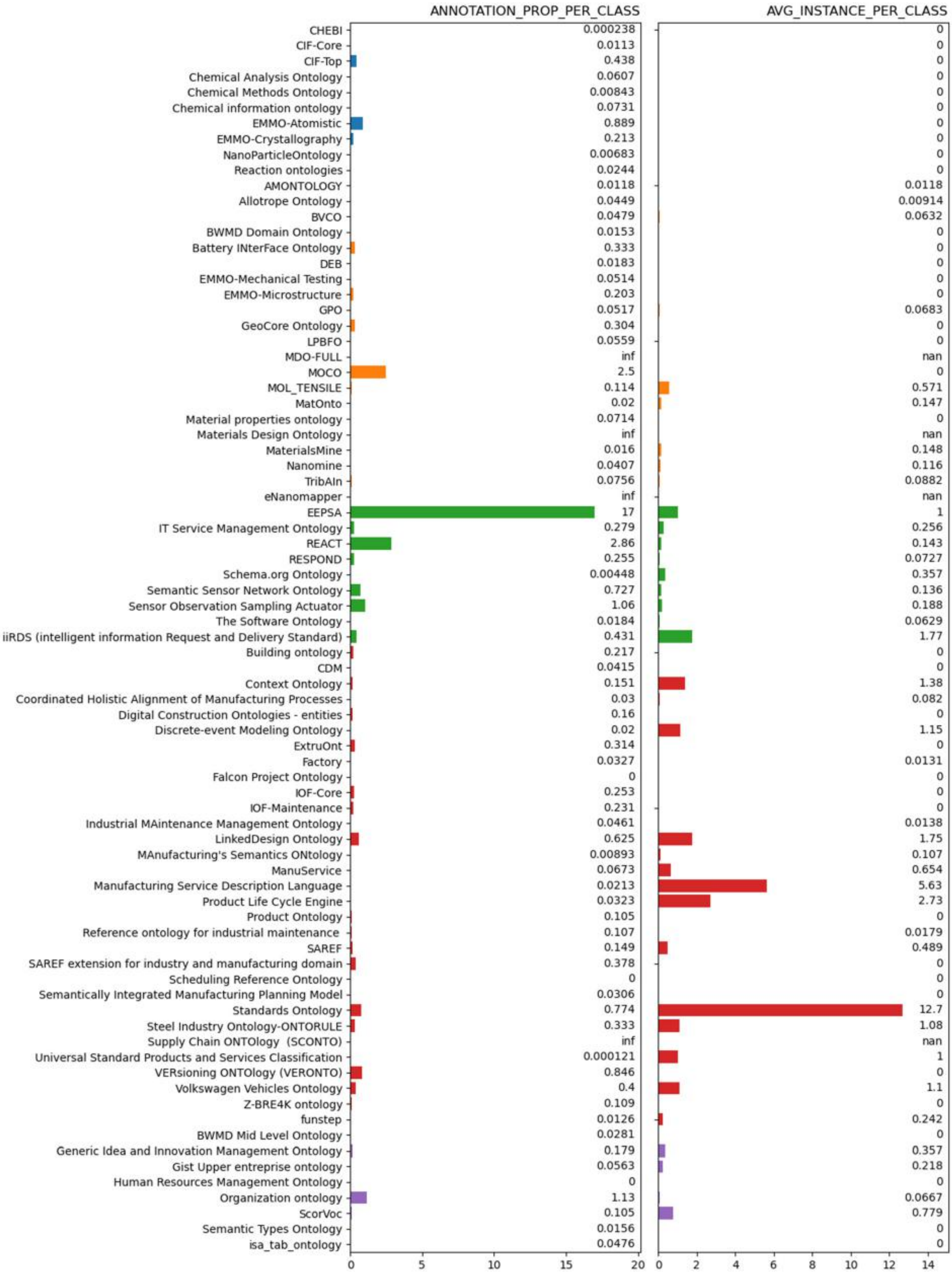

Figure 23 - Relative measures of the size of the ontologies per class: number of annotation properties per class and the number of instances per class. The colours correspond to domains according to Figure 19.

Figure 24 - Relative measures of the size of the ontologies per class: size of the TBox per class and the size of the ABox per class. The colours correspond to domains according to Figure 19.

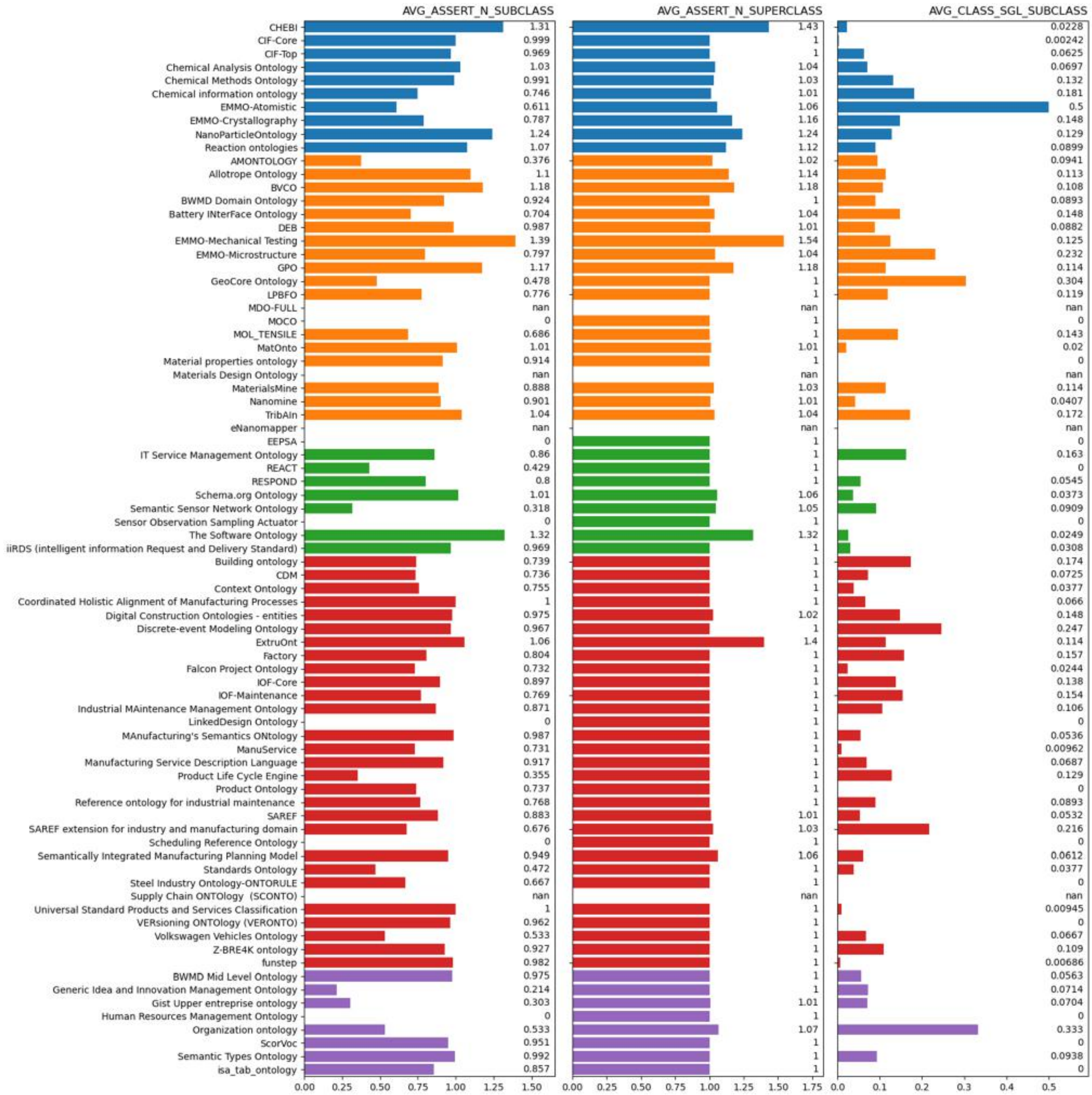

Figure 25 - Some numerical measures of the complexity of the ontologies: average number of asserted subclasses per class, average number of asserted superclasses per class and number of super-classes which have more than one subclass. The colours correspond to domains according to Figure 19.

Figure 26 - Some numerical measures of the complexity of the ontologies: number of classes with multiple inheritance and number of axioms with a complex right-hand side. The colours correspond to domains according to Figure 19

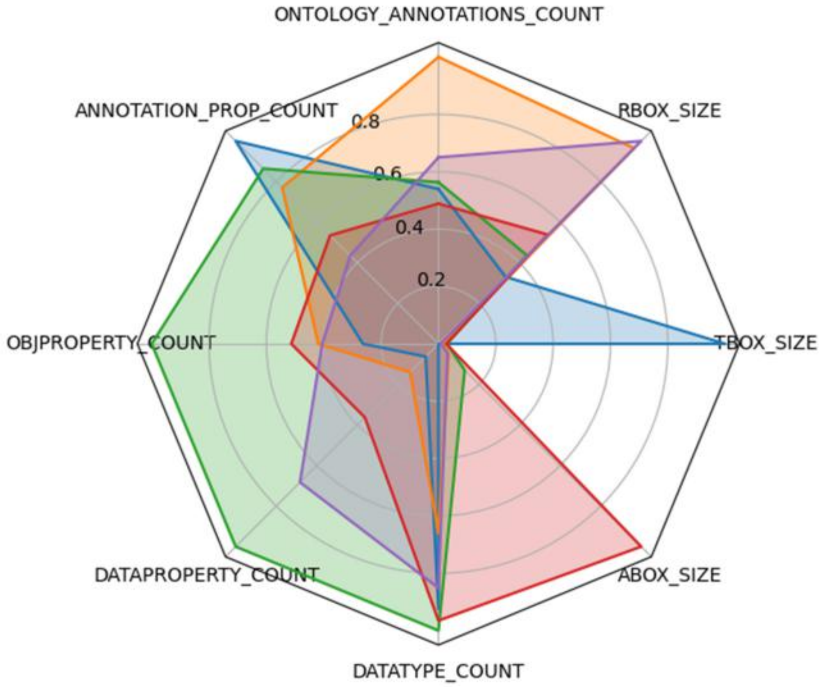

Figure 27 - The average within each domain (according to Figure 19) of some metric measures for the ontology size. See Table 6 for a description of each metric.

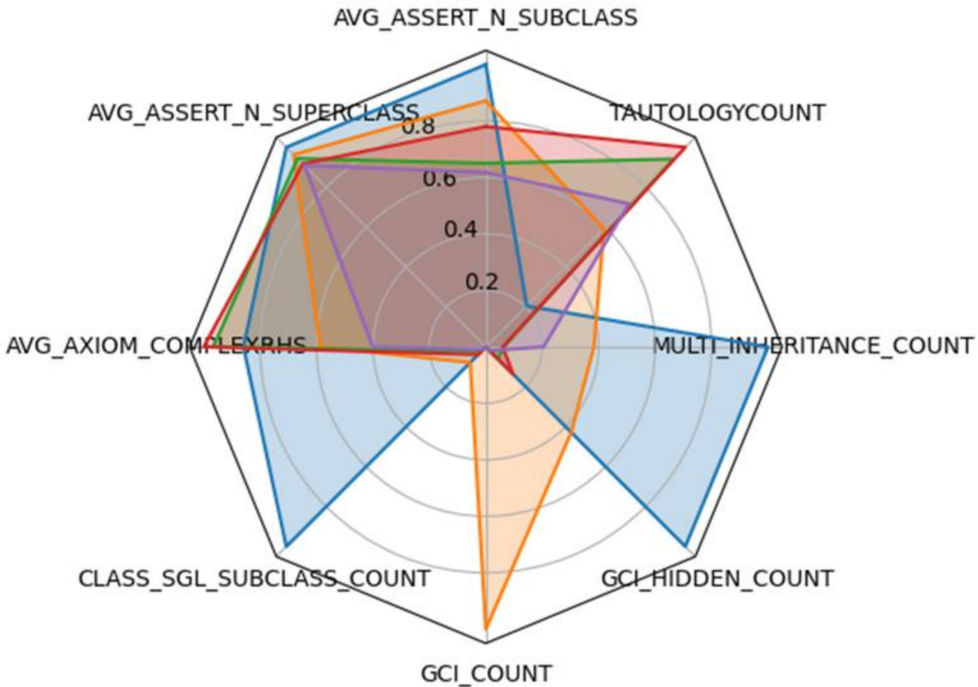

Figure 28 - The average within each domain (according to Figure 19) of some metric measures for the ontology complexity. See Table 6 for a description of each metric.

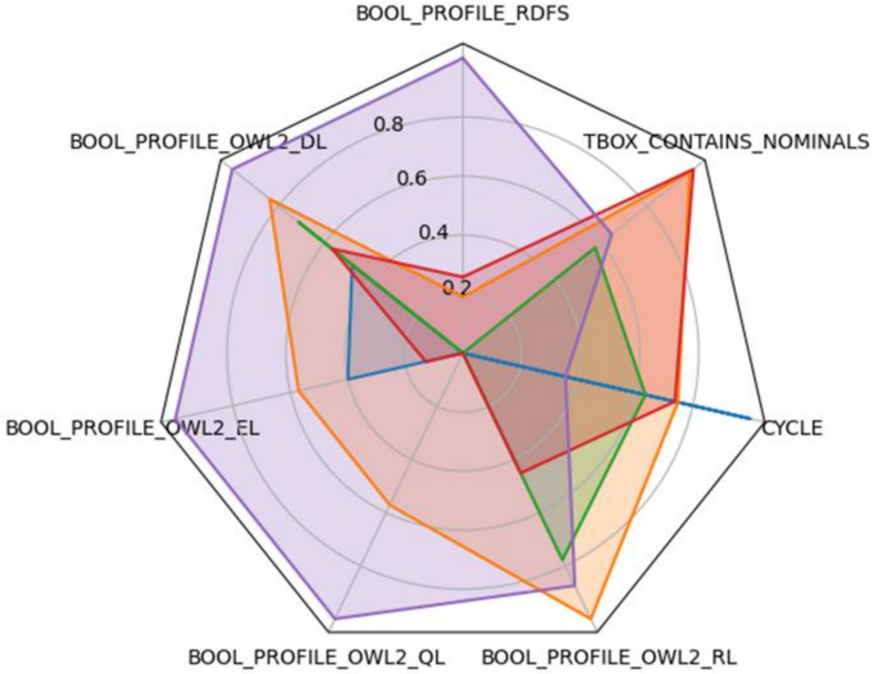

Figure 29 - The average within each domain (according Figure 19) of some Boolean metrics (where false corresponds to 0 and true to 1 in the calculation of the averages). See Table 6 for a description of each metric.

3.3.4.3 **Syntax**

Domain ontology models are found to be stored in one of the four types of syntaxes: OWL/XML, RDF/XML, OWL Functional, and Turtle. From the distribution of choices of these domain ontologies given in Figure 16, it can be observed that RDF/XML is the most popular format whereas OWL Functional is the least used.

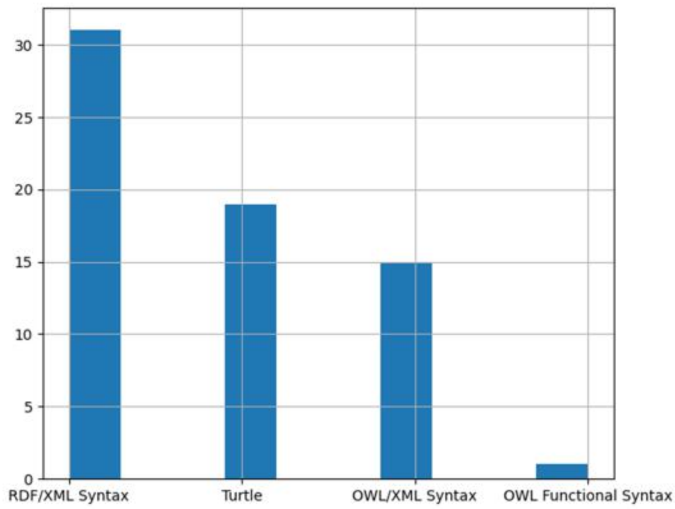

Figure 30 - Histogram with number of ontologies that uses the different syntaxes.

3.3.4.4 Expressivity

All the analysed ontologies are expressed using (languages based on) description logics with the expressivity of attributive language. This means that the ontologies enable one to express:

- Atomic negation (negation of concept names that do not appear on the left-hand side of axioms)
- Concept intersection
- Universal restrictions
- Limited existential quantification

However, they differ in supported features, listed in Table 11. The expressivity level of each of the analysed ontologies is shown in Table 12 in terms of the features listed in Table 11.

Table 11 - expressivity features

F	Functional properties, a special case of uniqueness quantification.
E	Full existential qualification (existential restrictions that have fillers other than Top).
U	Concept union.
C	Complex concept negation.
H	Role hierarchy (subproperties: rdfs:subPropertyOf).
R	Limited complex role inclusion axioms; reflexivity and irreflexivity; role disjointness.
O	Nominals. (Enumerated classes of object value restrictions: owl:oneOf, owl:hasValue).
I	Inverse properties.
N	Cardinality restrictions (owl:cardinality, owl:maxCardinality), a special case of counting quantification.
Q	Qualified cardinality restrictions (available in OWL 2, cardinality restrictions that have fillers other than Top).
D	Use of datatype properties, data values or data types.

Table 12 - The supported expressivity features for each of the analysed ontologies

NAME	F	E	U	C	H	R	O	I	N	Q	D
CHEBI		E						I			
CIF-Core				C		R		I		Q	D
CIF-Top				C		R		I		Q	D
Chemical Analysis Ontology				C	H						D
Chemical Methods Ontology				C	H						
Chemical information ontology	F			C		R		I			D
EMMO-Atomistic		E			H						D
EMMO-Crystallography				C	H			I		Q	D
NanoParticleOntology				C	H			I	N		D
Reaction ontologies				C	H						
Allotrope Ontology				C		R		I		Q	D
AMONTOLOGY		E									
BVCO				C		R	O	I		Q	D
BWMD Domain Ontology											
Battery INterFace Ontology		E									
DEB					H						D
EMMO-Mechanical Testing				C	H			I		Q	D
EMMO-Microstructure		E						I		Q	
GPO				C		R	O	I		Q	D
GeoCore Ontology				C		R		I		Q	
LPBFO				C	H						
MDO-FULL											
MOCO		E			H						D
MOL_TENSILE	F			C	H						D
MatOnto				C	H		O	I		Q	D
Material properties ontology										Q	D
Materials Design Ontology											
MaterialsMine		E					O				
Nanomine		E					O				
TribAIn				C		R		I		Q	D
eNanomapper											
EEPSA											
IT Service Management Ontology				C	H				N		D
REACT			U								D
RESPOND		E			H					Q	D
Semantic Sensor Network Ontology						R		I	N		D
Sensor Observation Sampling Actuator								I			D
The Software Ontology				C	H		O	I		Q	D
iiRDS (intelligent information Request and Delivery Standard)					H						D
Discrete-event Modeling Ontology				C			O			Q	D

Building ontology		E								Q	D
CDM				C	H		O	I		Q	D
Context Ontology	F			C				I			D
Coordinated Holistic Alignment of Manufacturing Processes				C		R		I		Q	D
Digital Construction Ontologies - entities	F			C	H			I			D
ExtruOnt		E						I		Q	
Factory				C			O			Q	D
Falcon Project Ontology				C	H					Q	D
IOF-Core				C							
IOF-Maintenance				C							
Industrial MAintenance Management Ontology				C	H						D
LinkedDesign Ontology	F							I			D
MAnufacturing's Semantics ONtology			U					I	N		D
ManuService											D
Manufacturing Service Description Language	F			C	H			I			
Product Life Cycle Engine	F			C	H		O	I			D
Product Ontology	F			C		R		I			
Reference ontology for industrial maintenance				C							
SAREF				C			O	I		Q	D
SAREF extension for industry and manufacturing domain				C	H			I		Q	D
Scheduling Reference Ontology		E								Q	
Semantically Integrated Manufacturing Planning Model				C	H					Q	D
Standards Ontology	F			C	H		O	I			D
Steel Industry Ontology-ONTORULE	F			C	H			I			D
Supply Chain ONTOlogy (SCONTO)											
Universal Standard Products and Services Classification											
VERsioning ONTOlogy (VERONTO)				C				I		Q	D
Volkswagen Vehicles Ontology			U		H						D
Z-BRE4K ontology											D
funstep				C	H			I		Q	D
ScorVoc				C	H		O				D
Semantic Types Ontology											
Human Resources Management Ontology											
Generic Idea and Innovation Management Ontology			U		H			I			D
Organisation ontology	F			C		R		I			D
Schema.org Ontology			U		H						D
isa_tab_ontology		E			H						D
BWMD Mid-Level Ontology				C	H						D
Gist Upper entreprise ontology				C		R	O	I		Q	D

3.4 FAIRness assessment

We have performed the assessment on 41 ontologies out of the 88 machine readable ontologies identified in our landscape analysis. We have then calculated three values: the FAIR Score that measures the percentage of mandatory recommendation fulfilled, the Global FAIR Score which evaluates the Semantic Artefact on the globality of the answers (i.e., including recommended and optional recommendations) and the FOOPS! Score that is the sum of sub principles FAIR scores divided by the total number of checks. The scores indicate how much a semantic artefact complies with the FAIR principles i.e., a score of 100% means that the semantic artefact complies with all the FAIR principles. Results are summarised in the table below.

Table 13 - Initial FAIR evaluation

Domain	Ontology Name	FAIR Score	Global FAIR Score	FOOPS! score
D1 Physics and Chemistry	Chemical Methods Ontology	50	46.2	39
	Reaction ontologies	50	46.2	48
	CHEBI	50	53.8	14
	Chemical Analysis Ontology	37.5	46.2	29
	Chemical information ontology	37.5	46.2	39
	NanoParticleOntology	25	46.2	38
	EMMO-Crystallography	25	38.5	44
	EMMO-Atomistic	25	38.5	31
	CIF Ontology	12.5	23.1	54
	Average (± STD)	34.7% (±13.7%)	42.7% (±8.7%)	37.3% (± 11.7%)
D2 Mechanical and Industrial Engineering	SAREF extension for industry and manufacturing domain	12.5	15.4	63
	Coordinated Holistic Alignment of Manufacturing Processes	25	38.5	20
	MAnufacturing's Semantics ONtology	12.5	23.1	04
	Semantically Integrated Manufacturing Planning Model	12.5	23.1	19

	Manufacturing Service Description Language	12.5	30.8	22
	ManuService	0	7.7	24
	Scheduling Reference Ontology	0	7.7	14
	Reference ontology for industrial maintenance	12.5	30.8	24
	funstep	12.5	23.1	25
	EMMO-mechanical-testing	25	38.5	43
	Factory	25	30.8	42
	Ontology for Simulation, Modelling, and Optimisation	37.5	38.5	35
	Industrial MAintenance Management Ontology	12.5	30.8	68
	Product Ontology	37.5	38.5	4
	Product Life Cycle Engine	0	7.7	4
	Volkswagen Vehicles Ontology	37.5	38.5	4
	SAREF	50	46.2	4
	IOF-Core	12.5	30.8	26
	Average (± STD)	18.8 % (±14.4%)	27.8 % (±11.8%)	24.6% (±19.29%)
D4 Materials Science and Engineering	eNanomapper	25	46.2	15
	EMMO-Mechanical Testing	25	38.5	44
	Battery INterFace Ontology	25	38.5	39
	EMMO-Microstructure	25	38.5	31
	BWMD Domain Ontology	25	30.8	04
	GeoCore Ontology	25	38.5	23
	Devices, Experimental scaffolds and Biomaterials Ontology	12.5	30.8	19
	MatOnto	12.5	30.8	35

	Allotrope Ontology	87.5	84.6	52
	TribAIn	25	30.8	27
	Average (± STD)	28.8% (±21.3%)	40.8% (±16.2%)	28.9% (±14.35%)
D5 Computer Science, Systems and Electrical Engineering	The Software Ontology	25	46.2	34
	Semantic Sensor Network Ontology	37.5	38.5	68
	Sensor Observation Sampling Actuator	37.5	38.5	67
	IT Service Management Ontology	25	30.8	4
	Average (± STD)	31.25% (±7.2%)	38.5% (±6.3)	43.25 (±30.6%)

The FAIRsFAIR analysis revealed some disparities between domains. Indeed D1-Physics and Chemistry is the domain with the highest FAIR Score on average. This is due to the fact that ontologies listed as part of the domain are following up some of the OBO Foundry recommendations. The "FAIRest" ontology is the Allotrope Ontology which is the only ontology tracking provenance using PROV.

In most of the cases, considering the Global FAIR Score improves the situation with respect to FAIRness. However, no ontologies passed the threshold of minimally FAIR (i.e., compliant with the mandatory FAIR recommendations).

Finally, it is important to mention that this analysis is ad-hoc to our project and should be the main theme for collaboration with FAIRsFAIR.

We observe that no ontology is totally compliant with the FAIR principles and only 6 ontologies over 44 have a score that is more or equal to 50%. FOOPS! results tend to be aligned with the FAIR score which is conservative (only consider the mandatory elements for FAIR): D1 (FAIR score 34.7%/ FOOPS! 37.3%), D2 (FAIR score 18.8%/ FOOPS! 24.6), D4 (FAIR score 28.8/ FOOPS! 28.9) and D5 (FAIR score 31.25%/ FOOPS! 43.25). This confirms that the tool considers the FAIRsFAIR recommendations as an evaluation reference. In addition, it considers other additional aspects such as tests related to version IRI resolving (VER2), protocol (HTTP1), documentation of labels (VOC3), documentation of definitions (VOC4), content negotiation (CN1), and vocabulary reuse (VOC1). We believe that these aspects are important and should be included in any semantic evaluation; for example, O'FAIRe includes all FOOPS! checks in its methodology that is composed of 62 FAIR questions.

Our analysis work demonstrates that existing approaches have similar aspects, however it also stresses the need to converge all these evaluation frameworks in order to come up with a unique methodology and scores.

3.5 Domain criteria (coverage, overlap, semantic gaps, usage, maturity)

Providing a detailed study on the coverage, gap, overlaps, usage, and maturity is extremely complex and requires us to have a deep understanding of all the ontologies collected. Furthermore, as we have experienced with our first attempt to categorise the collected ontologies by domain, classifying ontologies can be quite subjective. To provide a meaningful domain coverage and gap analysis requires domain expert contribution and validation as shown by the initial information derived from the community interactions (see Table 4, Table 5).

To providing a first evaluation of the domain coverage and overlap of ontologies within our dataset, we used two approaches: (1) we extended the categorisation of ontologies using the subdomains that are associated with the five high level domains initially used to analyse our dataset, and (2) we leveraged the mappings provided by the IndustryPortal and MatPortal to look at the degree of overlap between ontologies. The results of the categorisation are shown in Table 14 and the mapping matrices are shown in Figure 31 and Figure 32.

Through the classification exercise, we identified differences in the overall number of ontologies for each high level domain (i.e., D1, D2, D3, D4 and D5) in comparison to

Figure 11 for the following three reasons: (i) for

Figure 11 analysis, we allocated each ontology to one domain only, whereas in the more detailed analysis we also accounted for the relevance of certain ontologies to multiple domains, (ii) we included recent developments and additions, and (iii) not all ontologies in domains D2, D3, D4 and D5 were included. The mapping of ontologies to domains and sub-domains currently can only be done by experts inspecting the ontologies one by one, which is obviously very time consuming and not sustainable. It appears clearly that ontologies lack clear domain metadata. For a more automated and scalable analysis of domain applicability, at least the 'Domain' metadata should be required for ontologies.

Table 14 - Classifying ontologies by subdomains

Research area	Research subarea	DFG ref code	ERC ref code	Ontologies	# ontol-ogies
Physics and Chemistry (D1)	**Chemistry**	**31**			**17**
	Molecular Chemistry	301		ChemInf, CHEBI	2
	Chemical Solid State and Surface Research	302		BattInfo, EMMO-Crystallography, MDO, CIF, NPO, ENM, EMMO-Mechanical Testing, BWMD Domain Ontology, TribAIn, VIMMP, ChemInf, CHEBI	13
	Physical and Theoretical Chemistry	303		RXNO, ChemInf, CHEBI	3
	Analytical Chemistry, Method Development (Chemistry)	304		NPO, ENM, Chemical Methods Ontology, Chemical Analysis Ontology, DEB, Allotrope, ChemInf, CHEBI	8
	Biological Chemistry and Food Chemistry	305		NPO, ENM, DEB, ChemInf, CHEBI	5
	Polymer Research	306		DEB, ChemInf, CHEBI	3
	Physics	**32**			**8**
	Condensed Matter Physics	307		EMMO-Crystallography, MDO, CIF, TribAIn, VIMMP	5
	Optics, Quantum Optics and Physics of Atoms, Molecules and Plasmas	308			0
	Particles, Nuclei and Fields	309			0
	Statistical Physics, Soft Matter, Biological Physics, Nonlinear Dynamics	310		NPO, CHEBI, ChemInf	3

Mechanical and industrial engineering (D2)	**Mechanical and industrial engineering**	**41**			**27**
	Production technology	401		OFM, IOF-Core, ManuService, MSDL, MASON, SAREF4INMA, BWMD Domain Ontology, SCONTO, PRONTO, SIMPOM	10
	Mechanics and constructive mechanical engineering	402		ExtruOnt, VAR, PSS, LIDON, SOM, SCOR, QU4LITY-RMPFQ (Whirlpool, in progress),Di-Con, BIMERR	9
	Aerospace Engineering		PE8_ 1	QU4LITY-AIRBUS ontology, (in progress)	1
	Automotive Engineering				
	Industrial Maintenance			IOF-Maintenance, IMAMO, ROMAIN, Z-BRE4K	4
	Procurement, supplier and vendor engineering				
	Quality Control			QU4LITY-GFMS ontology (in progress), BOOST4.0-GFMS ontology	2
Thermal and Process Engineering (D3)	**Thermal Engineering/Process Engineering**	**42**			**4**
	Process Engineering, Technical Chemistry	403		BWMD Domain Ontology	1
	Heat energy technology, thermal machines, fluid mechanics	404		EEPSA, REACT, RESPOND	3
Materials Science and Engineering (D4)	**Materials Science and Engineering**	**43**			**14**
	Materials engineering	405		EMMO-Mechanical Testing, BWMD Domain Ontology, DEB, MatOnto, MMFO, NanoMine, Material properties ontology, TribAIn	8

	Materials Science	406		BattInfo, EMMO-Mechanical Testing, EMMO-Microstructure, BWMD Domain Ontology, DEB, MatOnto, MMFO, TribAIn, NPO, ENM, VIMMP	12
Computer science, systems and electrical engineering (D5)	**Computer science, systems and electrical engineering**	**44**			**9**
	Systems engineering	407		ADACOR, SOSA, SSN, SAREF, iiRDS	5
	Simulation engineering and modelling		PE7-3	MDO, VIMMP, EMMO-Atomistic, MMFO	4
	Networks (communication networks, sensor networks, networks of robots...)		PE7-8	iiRDS, SSN	2
	Biomedical Systems Technology	407-06			

This initial work revealed that several ontologies can be used for multiple subdomains. This is particularly true for Physics and Chemistry and Materials Science (for example, ENanoMapper and MatOnto are categorised in Materials Science but relate to the chemical aspects of Materials Science). We will pursue this analysis and collect expert validation in the course of the project.

Another issue we faced during this exercise is the discrepancy between the domain/subdomain classification we are using and the areas identified throughout our discussion with domain experts. In the Materials Science domain, focus areas defined by the domain experts are material properties, crystallography, statistical methods, characterisation, molecular materials, chemical kinetics, tribology, corrosion, powder materials, roles of chemicals and materials, materials for electronics, extensive physical properties, etc. These areas are on a finer grained level and are not actually reflected as such within our classification scheme. In the domain of industrial and manufacturing engineering, the identified focus areas are product, systems, supply chain, design, logistic, procurement, maintenance, and planning. These areas are difficult to map with our academic classification scheme. In the next iteration of this work, we should consider extending our classification to these specific focus areas, with possibly mappings to our initial academic classification.

Furthermore, this classification exercise provided us the means to identify some gaps within the semantic landscape. Some of these gaps have already been identified and discussed during expert workshops.

For instance, in the Materials Science domain, although some ontologies are identified for the focus areas mentioned above (e.g., material properties, crystallography, etc.), participants of the workshops emphasised the need for more extensive and granular models addressing these areas. Areas such as tribology, corrosion, and powder materials were identified by participants as an immediate focus for ontology development. Some of the areas for which no ontology has been found yet are in the sub-areas of chemicals and materials, materials for electronics, and extensive physical properties, to name but a few. There is also a general lack of ontologies that covers fundamental and application-specific physics and chemistry related topics, as evident from domain distribution statistics in

Figure 11.

Similarly, in the domain of industrial and manufacturing engineering, the most important areas are product, systems, supply chain, followed by design and planning. Some areas that are still waiting more coverage are logistic, procurement, maintenance, and planning. One critical aspect in this domain is contextual heterogeneity in the definitions of the terms as stakeholders from different phases of the life cycle view the manufacturing elements differently. Ontologies in process and thermal engineering are scarce as observed in in

Figure 11.

The feedback from the community, however, needs to be evaluated by a formal method. One method is to compare the concepts covered in different domain ontologies with a Gold Standard catalogue; several metrics related to coverage, overlap, and gaps are described in literature in this area (Dellschaft, 2006). Nonetheless, selecting suitable Gold Standard catalogues for each domain in focus is challenging and requires consensus among domain experts. Along with formal methods, more inputs from community and internal reviews will be conducted to understand the usage and maturity of these ontologies. These measures should be collected automatically through the use of ontology repositories.

It is already possible to investigate overlaps between ontologies by leveraging the ontology repositories mentioned in section 2.2.3 and their associated tools. In particular, the OntoPortal appliance includes Loom, a lexical mapping tool (Ghazvinian et al., 2009 Nov). Mappings are generated by comparing concept labels and synonyms. It also offers other means to identify mapping such as OBO Xref and URI matching (Salvadores et al., 2013).

We evaluated the numbers of lexical mappings between pairs of ontologies stored in IndustryPortal and in MatPortal. In Figure 31, the matrix contains the number of lexical mappings deduced between every two ontologies from IndustryPortal, which are mainly from the Industrial and Systems Engineering domain. In Figure 32, the matrix contains the number of lexical mappings deduced between every two ontologies from MatPortal, which are mainly from the Materials Science domain. Using the counts of mappings among the terms, every number in the matrices below expresses the overlap between two ontologies.

	Bonsai	CONT-ONT	DEMO	EMMO-MECH-TEST	ExtruOnt	FunStep	GI2MO	IMAMO	FALCON	GRACE	I40KG	MSDL	MASON	Manuservice	PRONTO	PSS	ROMAIN	SAREF4INMA	SRO	SCOR	SCOPRO	SWO	VAR	Z-brea4k
Bonsai			1	2	3		1	2			2							1	2				1	
CONT-ONT						1	1	2				1				2	1		1			1		1
DEMO	1			4				4		1	2	1	1	1		2	4	1	4	1	2	2	2	
EMMO-MECH-TEST	2		4		16	1		5		2	3	25	11	2		2	16	8	10	1	2	8	6	
ExtruOnt	3			16		2	1			1	5	27	2		1	1	20	8	3		2	3	3	1
FunStep	1	1		1	2						1	8	4	2		2	2		1	1	1	2		1
GI2MO	1	1			1						3			2								1		
IMAMO	2	2	4	5					1	5	4	7	3	1		6	11	5	10	4	5	3	4	2
FALCON					7			1		1		2	1	1	1	5							1	
GRACE			1	2	1			5	1		1	4	3	3	1	3	3	2	3		1		2	
I40KG	2	1	2		5	1	3	4		1		5	2	1		2	6	1	1	1		2	2	1
MSDL			1	25	27	8		7	2	4	5		12	9	2	43	231	4	6	1	4	28	7	8
MASON			1	11	2	4		3	1	3	2	12		3		3	6	2	2	1	3	1	3	2
Manuservice			1	2		2	2	1	1	3	1	9	3		1	2	2		4		1	1		
PRONTO					1				1	1		2		1		1						1		
PSS		2	2		1	2		6	5	3	2	43	3	2	1		41	2	3	1	4	12	2	17
ROMAIN		1	4	16	20	2		11		3	6	231	6	2		41		4	4	1	3	32	3	5
SAREF4INMA	1	1	1	8	8			5		2	1	4	2			2	4						2	
SRO	2		4	10	3	1		10		3	1	6	2	4		3	4			2	4	3	2	1
SCOR						1		4			1	1	1			1	1		2		38	3	1	
SCOPRO			1	2	2	1		5		1		4	3	1		4	3		4	38		2	1	1
SWO		1	2	8	3	2	1	3			2	28	1	1	1	12	32			3	2			3
VAR	1	1	2	6	3			4	1	2	2	7	3			2	2	2	3	1	1			
Z-brea4k					1	1		2			1	8	2			17	5		1		1	3		

1-9
10-19
20-50
>50

Figure 31 - Ontology overlaps expressed as the number of mappings between every pair of ontologies from IndustryPortal.

	AMONTOLOGY	BUILDMAT	BWMD_Domain	BWMD-Mid	CCO	DEB	EMMO	LPBFO	MOCO	MATONTO	MP-SCHM	MSEO	NMRRVOCAB	MDO-FULL	MM	MOL_TENSILE	VIMMP
AMONTOLOGY			2	1	2	2	1	2		3	2	6	15		6	1	2
BUILDMAT			3	1		1	2	1		2	2			3	2	1	7
BWMD_Domain	2	3		346	51	14	36	364		137	13	160	61	5	71	353	35
BWMD-Mid	1	1	346		50	3	34	346		131	12	922	2	5	62	346	34
CCO	2		51	50		8	22	51	1	34	8	3	9	2	87	50	27
DEB	2	1	14	3	8		9	3		10	1	20	79	3	31	3	1
EMMO	1	2	36	34	22	9		34	2	48	15	110	16	4	54	36	86
LPBFO	2	1	364	346	51	3	34			131	12	927	2	5	63	347	35
MOCO					1		2				2	2			7		2
MATONTO	3	2	137	131	34	10	48	131			62	117	46	6	222	132	24
MP-SCHM	2	2	13	12	8	1	15	12	2	62		11	26	7	28	12	21
MSEO	6		160	922	3	20	110	927	2	117	11		10	10	320	925	138
NMRRVOCAB	15		61	2	9	79	16	2		46	26	10		1	120	2	6
MDO-FULL		3	5	5	2	3	4	5		6	7	10	1		8	5	15
MM	6	2	71	62	87	31	54	63	7	222	28	320	120	8		66	71
MOL_TENSILE	1	1	353	346	50	3	36	347		132	12	925	2	5	66		34
VIMMP	2	7	35	34	27	1	86	35	2	24	21	138	6	15	71	34	

1-9
10-49
50-99
100-199
200-500
>500

Figure 32 - Ontology overlap expressed as number of mappings between every pair of ontologies from MatPortal.

Within the IndustryPortal, two ontologies have an extensive number of concept mappings, reflecting an important overlap: ROMAIN and MSDL (237). In the context of MatPortal, several ontologies have a large number of mappings (more than hundreds). In particular, the MSEO ontology overlaps very strongly with three ontologies: BMWD-Mid (922), LPBFO (927) and MOL-Tensile (925). Such discrepancies between the two repositories can be explained by the differences in scope. MatPortal is exclusively focused on Materials Science Ontologies while IndustryPortal has a much broader scope.

This approach allowed us to identify and quantify overlaps between ontologies. This information could be actually quite useful for domain experts and ontologists to refine and align these ontologies. In the meantime, this approach can be used to evaluate the domain coverage once the Gold Standard catalogues have been identified in the different domains.

4. Conclusions

In this document, we described our approach and the first set of results from the landscape analysis on domain ontologies for Materials Science and Manufacturing.

This analysis highlights the strong heterogeneity within and among the five main domains of interests in terms of number of ontologies, level of complexity, and alignment with TLOs. Based on these results it is possible to identify potential gaps. However, this first analysis has been carried out with a restricted scope. In order to further identify gaps in the semantic landscape, we are planning to investigate further the disparities between subdomains with the support of expert communities.

The analysis revealed the difficulty of working with multiple ontologies published in various places with different formats and syntax. To perform a more in-depth analysis, we need to automate the analysis. For this, we have initiated a collaboration with MatPortal[32] (an instance of OntoPortal for Materials Sciences) to store Materials Science-related ontologies into an ontology repository that would give access to both the metadata and the content of the various ontologies via a REST API. In addition, the IndustryPortal provides access to ontologies from other relevant domains. Using OntoPortal appliances also offers a large range of tools that will help us in deepening our analysis on a third set of criteria, more related to the domain coverage of the ontologies and their overlap. In the meantime, we would like to identify existing domain Gold Standards that would be useful to evaluate quantitatively the domain coverage of an ontology, to better identify the existing gaps, and to propose solutions for filling these gaps.

Finally, we have been considering in this analysis the level of FAIRness of the various ontologies we collected using two existing evaluation frameworks. To perform our analysis, we first analysed the FAIRsFAIR recommendations and proposed a unique and simple evaluation matrix which allowed us to measure the level of FAIRness for each of the machine-readable ontologies. The first results based on a subset of the ontologies (44 out of 88) distributed across the different domains highlight the low level of compliance to FAIR principles (below 50% FAIR). This analysis has been performed manually and is not scalable. We extended this analysis by using the recently published FOOPS! Web service (Garijo et al., 2021) which confirmed the results obtained with our scoring method based on

32 https://matportal.org/

FAIRsFAIR recommendations. We will in the future compare with another existing evaluation framework associated with AgroPortal (Amdouni et al., 2021). By comparing the results of the FAIRness evaluations, we are hoping to establish some convergence between the different evaluation frameworks, and ensure a more harmonised evaluation of FAIRness for ontologies. This work should be then integrated with the existing FAIR data evaluators such as FAIRSharing or F-UJI.

This document provides a snapshot of our landscape analysis. This work will be continuously updated during the course of the project as we encounter new ontologies and as we are refining our analysis. The final outcomes should be presented in a series of scientific publications.

5. References

Amdouni, Emna, Jonquet, Clement (2021) FAIR or FAIRer? An integrated quantitative FAIRness assessment grid for semantic resources and ontologies. https://hal-lirmm.ccsd.cnrs.fr/lirmm-03208544v2

Bahim, C., Casorrán-Amilburu, C., Dekkers, M., Herczog, E., Loozen, N., Repanas, K., Russell, K. and Stall, S., 2020. The FAIR Data Maturity Model: An Approach to Harmonise FAIR Assessments. *Data Science Journal*, 19(1), p.41. DOI: http://doi.org/10.5334/dsj-2020-041

Cox, S. J., Gonzalez-Beltran, A. N., Magagna, B., & Marinescu, M. C. (2021). Ten Simple Rules for making a vocabulary FAIR. *PLOS Computational Biology*, *17*(6), e1009041.

David, R., Mabile, L., Specht, A., Stryeck, S., Thomsen, M., Yahia, M., Jonquet, C., Dollé, L., Jacob, D., Bailo, D., Bravo, E., Gachet, S., Gunderman, H., Hollebecq, J.-E., Ioannidis, V., Le Bras, Y., Lerigoleur, E., Cambon-Thomsen, A. and Alliance – SHAring Reward and Credit (SHARC) Interest Group, T.R.D., 2020. FAIRness Literacy: The Achilles' Heel of Applying FAIR Principles. *Data Science Journal*, 19(1), p.32. DOI: http://doi.org/10.5334/dsj-2020-032

Dellschaft, K., & Staab, S. "On how to perform a gold standard based evaluation of ontology learning". *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *4273 LNCS*, 228–241 (2006). https://doi.org/10.1007/11926078_17

Dutta, B, Toulet, A, Emonet, V, Jonquet, C, 2017. 'New Generation Metadata Vocabulary for Ontology Description and Publication', in: Garoufallou, E, Virkus, S, Siatri, R, Koutsomiha, D (eds.), Metadata and Semantic Research, Communications in Computer and Information Science. Springer International Publishing, Cham, pp. 173-185. https://doi.org/10.1007/978-3-319-70863-8_17

Garijo, Daniel, and María Poveda-Villalón. "Best Practices for Implementing FAIR Vocabularies and Ontologies on the Web." *arXiv preprint arXiv:2003.13084* (2020).

Ghazvinian A, Noy NF, Musen MA. Creating mappings for ontologies in biomedicine: simple methods work. AMIA Annu Symp Proc. 2009 Nov 14;2009:198-202. PMID: 20351849; PMCID: PMC2815474.

Hugo, Wim, Le Franc, Yann, Coen, Gerard, Parland-von Essen, Jessica, & Bonino, Luiz. (2020). D2.5 FAIR Semantics Recommendations Second Iteration (Version 1.0), doi : 10.5281/zenodo.4314321.

Jackson, Rebecca C., Balhoff, James P., Douglass, Eric, Harris, Nomi L., Mungall, Christopher J., Overton, James A. (2019). ROBOT: A Tool for Automating Ontology Workflows, BMC Bioinformatics, 20, 407, doi: https://doi.org/10.1186/s12859-019-3002-3

Le Franc, Yann, Parland-von Essen, Jessica, Bonino, Luiz, Lehväslaiho, Heikki, Coen, Gerard, & Staiger, Christine. (2020). D2.2 FAIR Semantics: First recommendations (Version 1.0). FAIRsFAIR, doi: 10.5281/zenodo.3707985.

Matentzoglu, Nicolas, Malone, James, Mungall, Chris, Stevens, Robert (2018). MIRO: guidelines for minimum information for the reporting of an ontology, J. Biomed Semant, vol. 9, 6, doi: 10.1186/s13326-017-0172-7.

Poveda-Villalón, M., R. García-Castro, and A. Gómez-Pérez. "Building an ontology catalogue for smart cities." *Proceedings of the 10th European Conference on Product and Process Modelling, ECPPM*. 2014.

Poveda-Villalón, M., Espinoza-Arias, P., Garijo, D., & Corcho, O. (2020, September). Coming to Terms with FAIR Ontologies. In *International Conference on Knowledge Engineering and Knowledge Management* (pp. 255-270). Springer, Cham.

Salvadores, M., Alexander, P. R., Musen, M. A., & Noy, N. F. (2013). BioPortal as a Dataset of Linked Biomedical Ontologies and Terminologies in RDF. *Semantic web*, *4*(3), 277–284.

Wilkinson, Mark D., Dumontier, Michel, Sansone, Susanna-Assunta, Bonino da Silva Santos, Luiz Olavo, Prieto, Mario, Batista, Dominique, McQuilton, Peter, Kuhn, Tobias, Rocca-Serra, Philippe, Crosas, Mercè, Schultes, Erik (2019). Evaluating FAIR maturity through a scalable, automated, community-governed framework, Scientific Data, vol. 6, 174, doi: 10.1038/s41597-019-0184-5

6. Appendix

Table 15 - Basic topological metrics

#	NAME	AXIOM_COUNT	ONTOLOGY_ANNOTATIONS_COUNT	OBJPROPERTY_COUNT	DATAPROPERTY_COUNT	DATATYPE_COUNT	RBOX_SIZE	CLASS_COUNT
1	NanoParticleOntology	27764	5	65	16	3	70	1904
2	eNanomapper	0	20	0	0	0	0	0
3	EMMO-Mechanical Testing	1396	13	13	6	5	5	311
4	EMMO-Crystallography	359	2	7	1	2	2	61
5	Battery INterFace Ontology	97	3	1	0	1	0	27
6	EMMO-Microstructure	235	13	3	0	2	0	69
7	EMMO-Atomistic	64	11	3	1	2	1	18
8	The Software Ontology	15109	14	43	5	6	43	1845
9	BWMD Mid Level Ontology	1457	10	24	11	8	31	320
10	BWMD Domain Ontology	1803	11	0	0	1	0	459
11	GeoCore Ontology	116	3	12	0	1	12	23
12	Chemical Methods Ontology	27694	5	27	0	3	10	3084
13	Reaction ontologies	7093	6	14	0	3	8	901
14	SAREF extension for industry and manufacturing domain	336	16	28	11	3	19	37

15	Coordinated Holistic Alignment of Manufacturing Processes	14390	93	253	11	10	320	2001
16	MAnufacturing's Semantics ONtology	1255	2	26	17	6	10	224
17	Semantically Integrated Manufacturing Planning Model	335	1	26	7	3	10	98
18	Manufacturing Service Description Language	12816	3	116	0	3	92	422
19	ManuService	1512	0	33	183	9	0	104
20	Semantic Sensor Network Ontology	262	12	35	1	2	9	22
21	Sensor Observation Sampling Actuator	316	9	21	2	2	9	16
22	CHEBI	2560790	9	10	0	2	3	155779
23	Chemical Analysis Ontology	2646	9	22	1	4	13	445
24	Chemical information ontology	1644	17	52	6	7	56	342
25	Scheduling Reference Ontology	17	0	14	1	1	0	35
27	Gist Upper entreprise ontology	1710	3	109	23	4	60	142
28	ScorVoc	6475	29	5	249	5	248	285
29	Reference ontology for industrial maintenance	208	2	17	0	1	0	56
30	funstep	6393	2	78	95	6	54	875
31	MatOnto	5235	91	83	13	3	138	848

32	EMMO-mechanical-testing	1396	13	13	6	5	5	311
33	Allotrope Ontology	22057	65	311	26	9	473	2517
36	Factory	492	4	22	2	3	0	153
38	Industrial MAintenance Management Ontology	913	6	24	6	4	3	217
39	Product Ontology	353	1	31	0	1	21	38
40	Product Life Cycle Engine	2084	0	76	87	5	109	62
41	Volkswagen Vehicles Ontology	601	11	28	22	5	14	30
42	SAREF	735	14	30	6	5	3	94
43	IOF-Core	453	2	1	0	2	0	87
44	IOF-Maintenance	302	0	7	0	1	0	65
45	TribAIn	1063	3	59	6	3	85	238
46	Falcon Project Ontology	243	0	19	21	4	20	41
47	Steel Industry Ontology-ONTORULE	361	10	11	26	4	38	24
48	IT Service Management Ontology	573	10	41	8	6	25	43
49	Standards Ontology	6422	14	32	24	10	17	53
50	Human Resources Management Ontology	13	0	3	0	0	0	4
51	Universal Standard Products and Services Classification	99006	0	0	0	1	0	16505
52	Schema.org Ontology	8860	0	509	491	10	68	670

53	Organization ontology	650	34	34	3	2	20	15
54	LinkedDesign Ontology	524	2	19	63	4	6	8
55	Context Ontology	400	1	18	1	2	1	53
56	Semantic Types Ontology	510	0	0	0	2	0	128
57	Generic Idea and Innovation Management Ontology	576	4	58	28	4	39	28
58	Discrete-event Modeling Ontology	1287	0	50	26	6	0	150
59	Building ontology	410	8	15	17	7	0	46
61	Digital Construction Ontologies - entities	1060	11	126	18	5	99	81
62	ExtruOnt	324	7	18	0	2	1	35
63	iiRDS (intelligent information Request and Delivery Standard)	1699	0	41	21	4	60	65
64	isa_tab_ontology	246	0	18	24	3	40	21
66	Materials Design Ontology	7	11	0	0	0	0	0
67	Supply Chain ONTOlogy (SCONTO)	7	11	0	0	0	0	0
68	VERsioning ONTOlogy (VERONTO)	425	10	38	9	3	18	26
73	Z-BRE4K ontology	510	2	53	26	2	53	55
74	CIF-Core	11140	10	2	7	6	9	1238
75	CIF-Top	216	9	2	7	6	9	32

76	GPO	3792	20	47	3	7	131	542
77	BVCO	3964	21	50	3	7	132	585
78	Nanomine	812	8	1	0	1	0	172
79	Material properties ontology	549	7	13	8	5	0	140
80	EEPSA	18	12	0	0	0	0	1
81	RESPOND	363	9	6	6	5	1	55
82	REACT	175	12	1	23	2	0	7
83	CDM	1119	0	23	17	5	6	193
84	MOL_TENSILE	368	1	61	6	4	62	35
85	MOCO	14	7	2	2	2	2	2
86	AMONTOLOGY	219	0	5	0	1	0	85
87	LPBFO	511	7	2	0	1	1	143
88	MaterialsMine	2058	5	6	0	1	0	500
89	MDO-FULL	7	11	0	0	0	0	0
90	DEB	2138	3	12	109	3	68	601

Table 16 - Derived topological metrics

#	NAME	AVG_ASSERT_N_SUBCLASS	AVG_ASSERT_N_SUPERCLASS	CLASS_SGL_SUBCLASS_COUNT	MULTI_INHERITANCE_COUNT	AXIOM_COMPLEXRHS_COUNT	TBOX_SIZE_PER_CLASS	ABOX_SIZE_PER_CLASS	AVG_INSTANCE_PER_CLASS
1	NanoParticle Ontology	1.240021	1.240546	245	320	1591	8.548845	0	0
2	eNanomapper	nan	nan	0	0	0	nan	nan	nan
3	EMMO-Mechanical Testing	1.392283	1.540193	39	14	159	1.945338	0	0

4	EMMO-Crystallography	0.786885	1.163934	9	1	55	1.737705	0	0
5	Battery INterFace Ontology	0.703704	1.037037	4	0	5	0.888889	0	0
6	EMMO-Microstructure	0.797101	1.043478	16	0	22	1.115942	0	0
7	EMMO-Atomistic	0.611111	1.055556	9	0	6	1	0	0
8	The Software Ontology	1.320325	1.321409	46	7	2710	2.819512	0.072087	0.062872629
9	BWMD Mid Level Ontology	0.975	1	18	0	0	1.071875	0	0
10	BWMD Domain Ontology	0.923747	1	41	0	0	0.923747	0	0
11	GeoCore Ontology	0.478261	1	7	1	4	1	0	0
12	Chemical Methods Ontology	0.991245	1.031453	406	102	384	1.118677	0	0
13	Reaction ontologies	1.074362	1.119867	81	51	492	1.63374	0	0
14	SAREF extension for industry and manufacturing domain	0.675676	1.027027	8	0	42	2.27027	0	0
15	Coordinated Holistic Alignment of	0.9995	1.0005	132	11	289	1.301349	0.081959	0.08195902

	Manufacturing Processes								
16	MAnufacturing's Semantics ONtology	0.986607	1	12	0	26	1.522321	0.459821	0.107142857
17	Semantically Integrated Manufacturing Planning Model	0.94898	1.061224	6	0	46	1.77551	0	0
18	Manufacturing Service Description Language	0.917062	1	29	0	7	1.471564	10.27725	5.625592417
19	ManuService	0.730769	1	1	0	0	7.048077	0.673077	0.653846154
20	Semantic Sensor Network Ontology	0.318182	1.045455	2	0	73	3.772727	0.136364	0.136363636
21	Sensor Observation Sampling Actuator	0	1	0	0	0	0.0625	0.1875	0.1875
22	CHEBI	1.312237	1.43104	3550	0	85178	1.859025	0	0
23	Chemical Analysis Ontology	1.031461	1.040449	31	25	14	1.186517	0	0
24	Chemical information ontology	0.745614	1.011696	62	8	114	1.260234	0	0
25	Scheduling Reference Ontology	0	1	0	16	5	0.314286	0	0

27	Gist Upper entreprise ontology	0.302817	1.007042	10	80	113	2.584507	0.457746	0.218309859
28	ScorVoc	0.950877	1	0	0	0	1.891228	0.785965	0.778947368
29	Reference ontology for industrial maintenance	0.767857	1	5	3	34	1.392857	0.017857	0.017857143
30	funstep	0.981714	1	6	0	177	1.824	0.326857	0.242285714
31	MatOnto	1.005896	1.010613	17	215	610	2.113208	0.459906	0.14740566
32	EMMO-mechanical-testing	1.392283	1.540193	39	14	159	1.945338	0	0
33	Allotrope Ontology	1.096146	1.141438	285	138	868	1.55741	0.024235	0.009137863
36	Factory	0.803922	1	24	0	127	1.712418	0.013072	0.013071895
38	Industrial MAintenance Management Ontology	0.870968	1	23	6	101	1.396313	0.023041	0.013824885
39	Product Ontology	0.736842	1	0	0	13	3.394737	0	0
40	Product Life Cycle Engine	0.354839	1	8	22	22	7.129032	21	2.725806452
41	Volkswagen Vehicles Ontology	0.533333	1	2	0	0	3.466667	1.466667	1.1
42	SAREF	0.882979	1.010638	5	0	93	1.957447	0.489362	0.489361702
43	IOF-Core	0.896552	1	12	0	8	1.022989	0	0
44	IOF-Maintenance	0.769231	1	10	1	4	0.846154	0	0
45	TribAIn	1.037815	1.037815	41	0	59	1.457983	0.096639	0.088235294

46	Falcon Project Ontology	0.731707	1	1	0	1	3.365854	0	0
47	Steel Industry Ontology-ONTORULE	0.666667	1	0	0	0	4.25	1.458333	1.083333333
48	IT Service Management Ontology	0.860465	1	7	0	13	4.465116	0.255814	0.255813953
49	Standards Ontology	0.471698	1	2	0	7	2.056604	57.50943	12.69811321
50	Human Resources Management Ontology	0	1	0	0	0	1.5	0	0
51	Universal Standard Products and Services Classification	0.999697	1	156	0	0	0.999697	0.999697	0.999697061
52	Schema.org Ontology	1.014925	1.055224	25	18	0	3.973134	0.507463	0.356716418
53	Organization ontology	0.533333	1.066667	5	3	1	5.666667	0.066667	0.066666667
54	LinkedDesign Ontology	0	1	0	0	0	29.125	1.75	1.75
55	Context Ontology	0.754717	1	2	0	0	1.150943	1.566038	1.377358491
56	Semantic Types Ontology	0.992188	1	12	0	0	0.992188	0	0
57	Generic Idea and	0.214286	1	2	0	0	5.607143	0.607143	0.357142857

	Innovation Management Ontology								
58	Discrete-event Modeling Ontology	0.966667	1	37	0	185	2.966667	2.833333	1.146666667
59	Building ontology	0.73913	1	8	0	73	3.173913	0	0
61	Digital Construction Ontologies - entities	0.975309	1.024691	12	2	31	4.82716	0	0
62	ExtruOnt	1.057143	1.4	4	0	240	8	0	0
63	iiRDS (intelligent information Request and Delivery Standard)	0.969231	1	2	0	0	2.707692	7.646154	1.769230769
64	isa_tab_ontology	0.857143	1	0	0	22	5.714286	0	0
66	Materials Design Ontology	nan	nan	0	0	0	nan	nan	nan
67	Supply Chain ONTOlogy (SCONTO)	nan	nan	0	0	0	nan	nan	nan
68	VERsioning ONTOlogy (VERONTO)	0.961538	1	0	0	23	6.884615	0	0
73	Z-BRE4K ontology	0.927273	1	6	0	0	3.818182	0	0
74	CIF-Core	0.999192	1	3	0	1117	1.976575	0	0

75	CIF-Top	0.96875	1	2	0	18	1.75	0	0
76	GPO	1.173432	1.177122	62	10	317	1.950185	0.271218	0.068265683
77	BVCO	1.176068	1.179487	63	14	345	1.948718	0.251282	0.063247863
78	Nanomine	0.901163	1.005814	7	0	54	1.215116	0.116279	0.116279
79	Material properties ontology	0.914286	1	0	0	45	1.285714	0	0
80	EEPSA	0	1	0	0	0	0	1	1
81	RESPOND	0.8	1	3	4	5	1.054545	0.127273	0.072727
82	REACT	0.428571	1	0	0	0	3.857143	0.142857	0.142857
83	CDM	0.735751	1	14	33	155	1.709845	0.134715	0
84	MOL_TENSILE	0.685714	1	5	0	0	0.885714	0.571429	0.571429
85	MOCO	0	1	0	0	1	1.5	0	0
86	AMONTOLOGY	0.376471	1.023529	8	0	92	1.458824	0.011765	0.011765
87	LPBFO	0.776224	1	17	2	0	0.79021	0	0
88	MaterialsMine	0.888	1.03	57	0	138	1.164	0.188	0.148
89	MDO-FULL	0	0	0	0	0	0	0	0
90	DEB	0.986689	1.006656	53	14	73	1.733777	0	0