RESEARCH

Open Access

Ontologies4Cat: investigating the landscape of ontologies for catalysis research data management

Alexander S. Behr[1*], Hendrik Borgelt[1] and Norbert Kockmann[1]

Abstract

As scientific digitization advances it is imperative ensuring data is Findable, Accessible, Interoperable, and Reusable (FAIR) for machine-processable data. Ontologies play a vital role in enhancing data FAIRness by explicitly representing knowledge in a machine-understandable format. Research data in catalysis research often exhibits complexity and diversity, necessitating a respectively broad collection of ontologies. While ontology portals such as EBI OLS and BioPortal aid in ontology discovery, they lack deep classification, while quality metrics for ontology reusability and domains are absent for the domain of catalysis research. Thus, this work provides an approach for systematic collection of ontology metadata with focus on the catalysis research data value chain. By classifying ontologies by subdomains of catalysis research, the approach is offering efficient comparison across ontologies. Furthermore, a workflow and codebase is presented, facilitating representation of the metadata on GitHub. Finally, a method is presented to automatically map the classes contained in the ontologies of the metadata collection against each other, providing further insights on relatedness of the ontologies listed. The presented methodology is designed for its reusability, enabling its adaptation to other ontology collections or domains of knowledge. The ontology metadata taken up for this work and the code developed and described in this work are available in a GitHub repository at: https://github.com/nfdi4cat/Ontology-Overview-of-NFDI4Cat.

Keywords Ontology collection, Research data management, Catalysis, Semantic web, Ontology classification, Metadata

Introduction

As digitization of the scientific community advances, the need for FAIR (Findable, Accessible, Interoperable, Reusable) [1] data rises to ensure machine-processability of data. Enabling a higher data FAIRness, ontologies represent knowledge explicitly in a machine-understandable way. Ontologies are a collection of machine- and human-interpretable concepts and relations that represent entities and their interdependence in a specific domain [1, 2]. Furthermore, research data occurring in the field of catalysis research often is complex and diverse. Thus, further ontology development and insights for the catalysis research domain are needed [3, 4].

To enhance semantic interoperability and compliance with existing ontologies, a collection of ontologies and semantic artefacts was created with importance to the data value chain of catalysis research. In addition, these ontologies and semantic artefacts were classified regarding their respective subdomains of research shaping the landscape of ontologies for catalysis research [5].

*Correspondence:
Alexander S. Behr
alexander.behr@tu-dortmund.de
[1] Laboratory of Equipment Design, Faculty of Biochemical and Chemical Engineering, TU-Dortmund University, Emil-Figge-Strasse 68, 44139 Dortmund, NRW, Germany

Domain and ontology experts browse and find the proper ontologies for their respective use usually with the help of portals, such as EBI OLS [6] and BioPortal [7]. However, these portals do not provide deep and classification on the ontologies. For example, the knowledge domains represented by an ontology are not covered by any of the two services. This means, that a pre-classification of the ontologies regarding the covered knowledge domain(s) is missing, yet desired. While there are general quantity metrics available, such as, e.g., number of contained classes in an ontology, quality metrics regarding the (re-)usability of ontologies are missing, such as if and which reasoning machine works on the ontology.

In 2022, Strömert et al. [8] screened 22 ontologies representing concepts for research data management in chemistry, also focussing on the (re-)usability, in the context of the NFDI4Chem project. As the authors wants to foster FAIR research data management in chemistry, the publication provides an overview on existing chemistry ontologies, evaluating them against criteria derived from the FAIR principles. Criteria for the evaluation of the ontologies include findability and accessibility of the ontologies, the modularity and alignment to an upper level ontology, as well as the availability of license information. For ontologies to be in scope of the survey, they need to contain a defined set of chemical sub-disciplines, made by domain experts, published and maintained in a FAIR way as well as being used in established applications. Furthermore, advantages and disadvantages of each ontology are discussed and areas of further improvement are highlighted, such as alignment with other ontologies for improved usefulness of the ontologies for FAIR data management [8].

Further classification of ontologies is done in the AIOTI Ontology Landscape Report [9]. Here, a comprehensive and thorough analysis of existing ontologies in the field of Internet of Things (IoT) and Artificial Intelligence (AI) takes place. The report also has a focus of the potential for interoperability and standardization of the ontologies. A total of 31 ontologies are analyzed and an overview of this is given with direct links to more detailed review documents for each ontology. Moreover, ontologies are evaluated on, among others, their functionalities, level of expressivity, and the technology readiness level. However, the report mainly lists ontologies from the domains healthcare, smart cities, energy, agriculture, and transportation. While the methodology and the metadata of the ontologies presented in [9] are well posed and the surveys are conducted thoroughly, most of the ontologies investigated possess minimal or negligible intersection with the domain of catalysis.

Another approach in collecting ontology metadata is realized by the OBO Foundry Dashboard [10], providing insights and metrics of ontologies within the Open Biological and Biomedical Ontologies (OBO) Foundry [11]. The dashboard provides a range of insights and metadata related to the ontologies, such as reuse of the respective ontologies in other ontologies. However, the dashboard only provides ontologies contained within the OBO library and the metrics only focus on (re-)usage related factors. Furthermore, the dashboard does not contain information on the respective scopes of the ontologies with regard to knowledge domains.

The initial listing and the classification of the ontologies and semantic artefacts presented in [5] was not sufficient. In addition, some of these ontologies are not easily reusable and do not provide proper documentation, as also denoted in [8]. As the work presented in [8] provides an overview of ontologies but with regards to the chemical domain of research, the general approach is used as inspiration for this work and is extended accordingly. Additionally, methods regarding the summarizing of ontologies as presented in [9] are considered in this work. Thus, this work presents a reiteration of the initial ontology landscape for catalysis research [5], focusing also on the classification of ontologies and other metadata important for the application of the ontologies listed. While focussing on comprehensibility of the resulting classification of ontologies, the workflow and software is developed to be as reusable as possible, to enable other domains for such ontology classification. Furthermore, a method is developed for a “lightweight” mapping of ontology classes and applied to the investigated ontologies.

Methods

Ontology metadata collection for domain relevance of ontologies

To identify suitable ontologies, ontologies listed in the EBI OLS [6] and BioPortal [7] are screened by look-up of classes and keywords by domain and ontology experts. Additionally, the ontologies listed in [5] are considered where suitable together with the overview on the landscape of ontologies in chemistry [8]. The ontology survey is conducted with the help of an intuitively designed spreadsheet template in Microsoft Excel to simplify access and handling of the ontology collection, capturing the relevant information on each ontology. This collection of ontology metadata with focus on the domain relevance is conducted for each ontology. Thus, for each ontology such a template is filled in consisting of six sections listed in Table 1 along the content included in each section.

In the following subchapters, the six spreadsheet sections that were evaluated for the ontologies are

Table 1 Classification scheme of the ontologies

Section	Content
General information on the ontology	Ontology name; alternative names; ontology acronym; creator(s) and issuing organization; kind of organizational structure
References	Organizational website; persistent URI of ontology file; link to documentation; link to version directory; additional links
Ontology modeling and availability	Provided ontology formats (ttl, owl,...); degree of inference and composition (inferred, non-inferred, compacted,...); license; working reasoners; shortest reasoning time; alignment with top level ontology; ontology imports; prefixes used; class annotation types
Classification of contained domains of interest	Biocatalysis; heterogenous catalysis; homogenous catalysis; photocatalysis; electrocatalysis; chemical substance modeling; material modeling; process modeling; synthesis data; operando data; performance data; characterisation data; heat, transport and kinetic data; process design; energy and cost data; top level ontology
Ontology characteristics	Axioms; logical axiom count; declaration; class count; object property count; data property count; individual count; annotation property count
Comments	Any additional comments or remarks on topics not covered by the other topics

Information regarding the six spreadsheet sections is gathered for each ontology to classify the ontologies regarding the content of each spreadsheet section

elucidated. Furthermore, the respective entries on the content listed in Table 1 are explained.

Spreadsheet section: general information on the ontology

To collect general information on the ontology, the ontology name and alternative names are gathered, if they exist. As ontologies often are referred to via their acronym, and stored using the acronym, it is also taken into the metadata. To get insights, whether or not the ontology still is maintained and if the ontology was developed by a consortium or a single person, metadata on the creator(s) and issuing organization is taken into account as well as the kind of organizational structure that developed the ontology.

Spreadsheet section: references

The section References of the scheme is intended to collect predominantly Uniform Resource Locators (URLs) for more information on the ontology. This encompasses the website of the organization that issues the ontology to get easy access on eventual updates or new releases of the ontology. Furthermore, the persistent URI of the ontology file is provided, which might be one of the most important metadata collected, as this allows for automated read in and manipulation of the ontology file with, e.g., Python. Some of the established Ontology Lookup services do not use persistent URIs and as such often reference older or deprecated versions, content pages and respective information. Further collected information are the links to the documentation of the ontology as well as to a version directory, if they exist. To account for further web resources on the ontology, such as web links to describing publications and the like, a metadata field for additional links is provided.

Spreadsheet section: ontology modeling and availability

Another scope of the metadata collection is to account for the availability and modeling depth of the ontologies. As ontologies can exist in different formats, the first metadata field deals with the available formats of the ontology files, such as Terse RDF Triple Language (TTL) [12] or Web Ontology Language (OWL) [13]. Furthermore, reasoning of the ontologies is an important aspect to obtain explicit knowledge from the otherwise only implicit defined knowledge contained in relationships within the ontology. The degree of inference and composition is collected for information on the availability of already inferred versions of the ontology, and whether compacted versions of the ontology are available. As ontologies are often setup in a modular way consisting of multiple sub-ontologies, a compacted version of an ontology contains all modules merged into a single file with no imports from other ontology files. Especially with focus on the reusability of ontologies, information on the respective license is important. Furthermore, the inference machines (reasoner) used on ontologies differ slightly in their execution, leading into some reasoners not working properly on some ontologies, implying a violation of the implied logic. Thus, the metadata field working reasoners captures the names of the reasoners that work on the ontology and contain empty entries, if there is no reasoner found working on the ontology. Only if the reasoning works for at least one reasoner, the shortest reasoning time can be captured as additional metadata field. Additionally, information about alignment with top level ontologies, such as the Basic Formal Ontology [14], ontologies that are imported via import statements into the ontology are captured by respective metadata fields. The collection of prefixes and class annotation types used in the ontology allows to directly get information on which prefixes are used in the ontology for, e.g., labels of classes.

Spreadsheet section: classification of contained domains of interest

As the reuse of an ontology not only depends on its availability and technical circumstances, such as licensing information, a classification with regards to the domains of interest contained within an ontology is also taken into account by the metadata fields. Here, the fields of catalysis research as listed in Table 1 are listed as metadata fields. To decide, whether an ontology enables for classification in the respective subdomain of catalysis research, the entries of the respective metadata fields are filled by screening the class hierarchy of the respective ontology. Where feasible, the textual definitions and annotations of the classes are also considered for the decision on domain relatedness. Furthermore, the available documentation and description of the ontology are taken into account. If many classes contained in the ontology are subject to a subdomain of catalysis research, the entry of the respective subdomain is set to contained, if close to no concepts or no concepts are contained to represent the subdomain, the entry is set to missing. Another classification is done by the rather subjective related:broader and related:narrower concepts, which try to indicate how well the subdomain in question is represented within the ontology.

Spreadsheet section: ontology characteristics

Further metadata on the ontology is captured by the section of ontology characteristics, which are aligned with the ontology metrics field of an ontology within Protégé [15]. Here, the number of axioms, logical axiom count, and declaration axioms count provide an idea of the semantic complexity of the ontology. Additionally, the class count, object property count, and data property count are provided to give a more thorough idea on the complexity of the ontology as well as the size. As individuals can provide for examples of the use of classes described within an ontology, the number of individuals also is included. Finally, the annotation property count gives the number of already available annotation properties within the ontology.

Spreadsheet section: comments

Any additional comments or remarks on topics not covered by the other topics are gathered within the metadata field of the comments section. This is important, as there might be, e.g., remarks on some of the gathered metadata fields of the other sections, additional information on the metadata collected or additional information about (re-)usability of the ontology in other software.

Documentation of the recorded metadata

To get the collected metadata of the ontologies into a more representable form whilst also ensuring machine readability of the data, the following workflow is set up. In a first step, the ontology metadata taken up in a Microsoft Excel file, structured as described in Sect. "Ontology metadata collection for domain relevance of ontologies" is converted to Markdown files. Markdown is a lightweight markup language favored for its simplicity and readability and can be rendered automatically on platforms such as GitHub. This provides users with well-formatted and easily accessible content especially used in *readme* files.

It also allows for linking between different Markdown documents, thus interconnecting the metadata aspects. Furthermore, the metadata is converted and stored as JavaScript Object Notation (JSON) files to enable for machine-readability.

The code utilizes the Python Pandas library [16] to ingest the ontology metadata as provided in the Microsoft Excel file. A JSON template is read in that acts as a blueprint for organizing the ontology metadata. The code extracts the information from the Microsoft Excel sheets and integrates it in a Python dictionary setup in the manner of the JSON template. This operation results in a comprehensive representation of each ontology. The resulting JSON data is saved into distinct files, each named after the respective ontology. Markdown files are generated on basis of the Python dictionary, which serve as easy accessible description of each ontology. The main *readme* Markdown file of the repository is also updated. A section within the file is created that contains links to the individual ontology Markdown files. These links and Markdown files are dynamically generated based on the ontologies characterized in the Microsoft Excel file.

Furthermore, the code undertakes an analysis to determine the suitability of ontologies for specific domains of interest of the catalysis research domain as listed in Table 1. This involves classifying ontologies based on their relationships (missing, contained, related:narrower, related:broader) to these domains as described in Sect. "Spreadsheet section: classification of contained domains of interest". To provide visual representations of the ontology relationships, radar plots are generated to categorize ontologies based on the relatedness. This also offers an quick and intuitive way to grasp the connections between ontologies and domains of interest. In addition to radar plots, the code creates Markdown tables summarizing the ontology relationships. These tables are subsequently incorporated into the main *readme* file and accessible to visualize the respective ontologies for each research domain, clustered by the respective relatedness. For each specific ontology, the code produces radar plots tailored to represent its relationships with the domains of interest. Thus, the entire ontology metadata processing workflow is executed, generating the structured documentation of the ontology metadata.

Ontology mappings

By collecting, among others, not only the relatedness of the ontologies to the respective domains of research, but also the URLs of the ontology raw-files, the collection can be used to automate various tasks regarding ontology analysis. However, it's worth noting that not all ontologies are provided in the standard OWL syntax, which is the format of ontology files needed for read in using the owlready2 [17] Python package. Thus, an automated conversion takes place where necessary and possible of the ontologies from TTL to the OWL syntax using ROBOT [18]. This ensures that the ontologies are properly loaded into Python with owlready2, allowing for comparison of classes across different ontologies.

The comparison functionality can be viewed as a preliminary mapping of ontology classes, offering an initial assessment of compatibility and relatedness between pairs of ontologies. This provides valuable insights into the potential overlaps and synergies between different knowledge representations. Classes are considered identical if they share the same Internationalized Resource Identifier (IRI), indicating a direct correspondence. Furthermore, classes within ontologies can be named by the annotations name, rdfs:label, rdfs:prefLabel, or skos:altLabel. This presents a challenge in comparing classes, as there can be multiple potential matches not covered by just comparing one way of class annotation against each other. For example, a class might be named via rdfs:label in one ontology, while the same name could be categorized as a rdfs:prefLabel in another. This necessitates a flexible approach to matching ontology classes.

For this, a systematic procedure is followed to streamline ontology comparison by iteration through the ontologies listed in the metadata collection. Using the owlready2 package, a list of classes within the ontologies in Python is retrieved. For each class, compiles a dictionary is compiled that includes the corresponding Internationalized Resource Identifiers (IRIs), along with the associated class attributes name, rdfs:label, rdfs:prefLabel, and skos:altLabel. In cases where one of the attributes is not available, the respective value is filled with none.

This process is repeated for each ontology, and the resulting dictionaries for each ontology are consolidated in an overarching dictionary. This approach eliminates the need to call upon the ontologies each time a comparison is required, accelerating the class comparison in contrast to the alternative of loading each pair of ontologies with owlready2 for every comparison. Then, for each pair of ontologies, the overarched dictionary is called to check for similar entries in the pair of sub-dictionaries for each ontology. To avoid redundancy, matches based on the IRI are searched for first. If a match is found, the class is excluded from further searches for similar name, rdfs:label, rdfs:prefLabel, and skos:altLabel. With this, the method is gathering both the total number of mapped classes and a detailed list of the respective classes from each ontology, facilitating a seamless comparison of the resulting mappings.

While it gives a hint of a potential mapping between classes, the decision only takes place based on same annotation (such as the name of the class). This neglects potential class definition strings or their embedding in the semantic web. Thus, the method presented serves as a “lightweight” approach to mapping ontology classes, offering an initial assessment of their compatibility and interconnection.

Moreover, the code extends its utility beyond automated comparison. Users have the option to apply the same method to their own lists of concepts, enabling the identification of the most suitable domain ontologies for a specific research area related to catalysis research, as defined by the concepts provided by the user. Another notable feature of the method is its capability to utilize the latest versions of ontologies for comparison. This functionality addresses a limitation of existing tools, such as BioPortal, which does not provide automated comparison of the most up-to-date ontology versions at the time of this work's publication.

Results and discussion

As described in Sect. "Introduction", the listings and different sources of ontology collections are screened and the metadata of the ontologies are collected using a template as described in Sect. "Ontology Metadata collection for domain relevance of ontologies". With this, a total of 30 ontologies are selected for further screening. They were obtained by search for domain specific keywords using OLS and BioPortal as well as regular web search engines and/or imported ontology classes within the found ontologies. Out of these, a decision was made to exclude seven of these ontologies. This decision was based on issues uncovered during the screening process, primarily related to the availability and accessibility of ontology files. For instance, some of the ISO 15,926 ontology files are proprietary, making it not freely accessible for further metadata collection. Additionally, upon closer examination, some ontologies exhibit lower domain relevance, are outdated and/or no version of the ontology file could be found. Table 2 lists the 30 ontologies, while the seven neglected ontologies are denoted

Table 2 Listing of the ontologies selected for further screening in this work

AFO [19]	CHMO [20]	ISO 15926* [21]	OBI [22]	OSMO [23]
BAO [24]	CIF [25]	ISO 15926-14* [26]	OFM* [27]	PIMS-II* [28, 29]
BFO [14]	DOLCE* [30]	M3 [31]	OM [32]	REX [33]
CAO [34]	EDAM [35]	M4I [36]	OntoCAPE [37]	RXNO [38]
ChEBI [39]	EMMO [40]	MOP [41]	OntoCompChem* [42]	SBO [43]
CHEMINF [44]	ENVO [45]	MS [46]	OntoKin* [47]	VIMMP [48]

Entries with an asterisk(*) denote the onologies not concerned any further for the modelling of catalysis research

with an asterisk(*) entries resulting in a total of 23 ontologies investigated in the metadata collection.

Investigating the ontology metadata

The analysis of the metadata of the ontologies reveals several observations regarding their expressivity in the research domains of catalysis research. First, it becomes evident that the domain of catalysis research itself lacks of uniformity, and the existing ontologies fall short in providing thorough descriptions of the research domains. Only four subdomains (Characterisation Data, Chemical Substance Modelling, Material Modelling, and Process Modelling) are described in multiple ontologies. In contrast, most subdomains lack a large number of matching ontologies. The domains Biocatalysis, Operando Data, Performance Data, and Process Design, Energy and Cost Data have only a single ontology that has the concepts of the respective domain contained. In addition, the domains Heterogeneous Catalysis, Homogeneous Catalysis, Photocatalysis, Electrocatalysis, Synthesis Data, and Heat, Transport and Kinetic Data have no dedicated ontology. This makes four of the 14 domains of catalysis research deemed as contained in multiple ontologies, while additional four domains are at least contained in a single ontology, while six domains are not described by any ontology.

Figure 1 shows the number of ontologies related to the respective domain of catalysis research in a radar plot, using the Python Plotly library [49].

Red denoted are the number of ontologies that are at least related:broader, yellow depicts the number of ontologies that are at least related:narrower, and green depicts the number of ontologies that have the respective domain of catalysis research contained. The specific fields of catalysis are denoted in blue, while the fields more directed to modelling are colored purple. Fields regarding general catalytic data are written in black.

Using multiple ontologies to model a domain of knowledge of catalysis research enables a more nuanced representation of diverse subdomains as more concepts might be contained in the respective ontologies to model the domain. Additionally, it is important to highlight that some of the listed ontologies pose challenges in terms of reasoning, as neither HermiT [50] nor FaCT++ [51] were able to effectively process them. Furthermore, the expressivity of some ontologies should be questioned, as the number of classes diverge widely in the different ontologies. This indicates a need for further refinement and development in this area of ontology engineering to ensure robust and comprehensive coverage within the field of catalysis sciences.

To facilitate documentation and ease access, the content of the ontology metadata listed in the Microsoft Excel file is used to automatically generate Markdown files that contain simplified, text-based formatting instructions and can be rendered similarly to Hypertext Markup Language (HTML). Rendering the Markdown files in GitHub provides a comprehensive and interactive overview of each ontology, making it easier for researchers to assess the suitability of an ontology for their research needs. Thus, the structure of the repository [52] is outlined below. The landing page of the repository shows the main *readme* file, which is formatted in Markdown syntax. It provides an overview on the whole ontology metadata collection, such as the radar plot shown in Fig. 1. Furthermore, a listing of the 23 ontologies is provided, containing the abbreviation and the full name of the ontologies. As Markdown allows for interlinking of files, the abbreviations of the ontologies link to separate Markdown files. Beside the general metadata of the ontology, a radar plot is contained in these Markdown files, showing the categorization of the respective ontology into the 14 domains of catalysis research similar to the one presented in Fig. 1. An excerpt of the rendering of the main *readme* Markdown-file and excerpts of the rendered page for the ChEBI ontology are depicted exemplarily in Fig. 2.

The data is also exported via Python to JSON files, by assigning key-value pairs of the respective metadata fields. This increases the machine-readability of the results, which eases further use of the data, such that other software can easily read out the metadata of the ontologies. Finally, the main *readme* Markdown-file also contains an overview on the mappings generated for pairs of ontologies. This overview and the respective results are described in the following section.

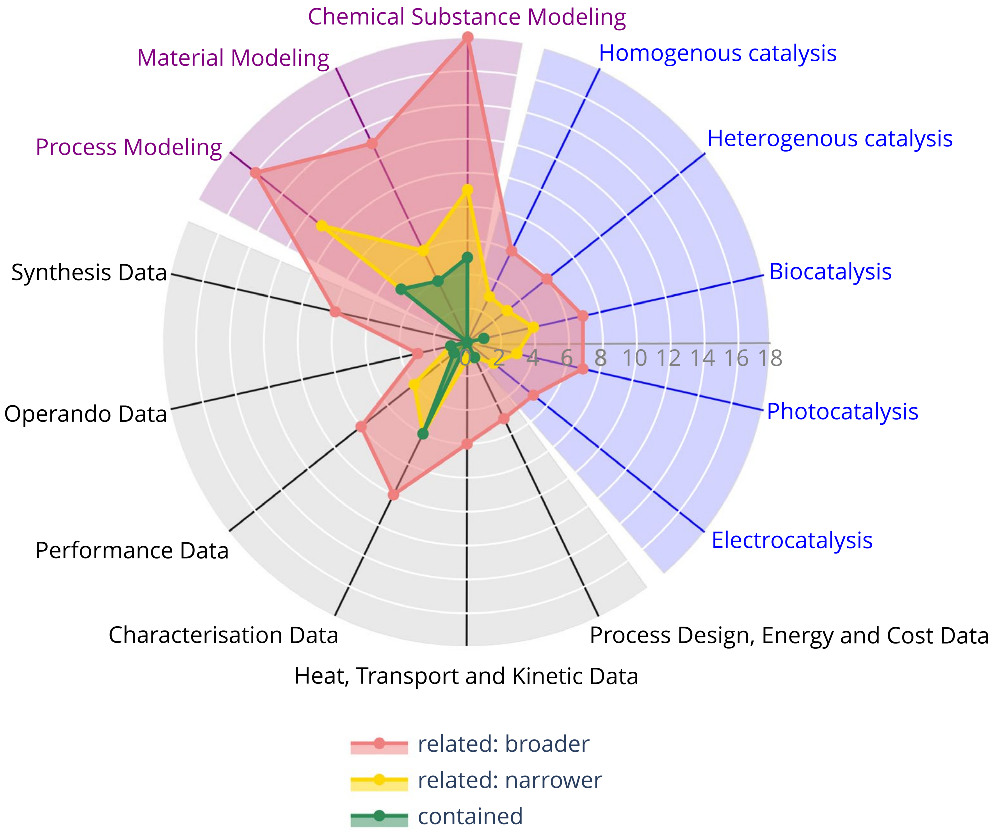
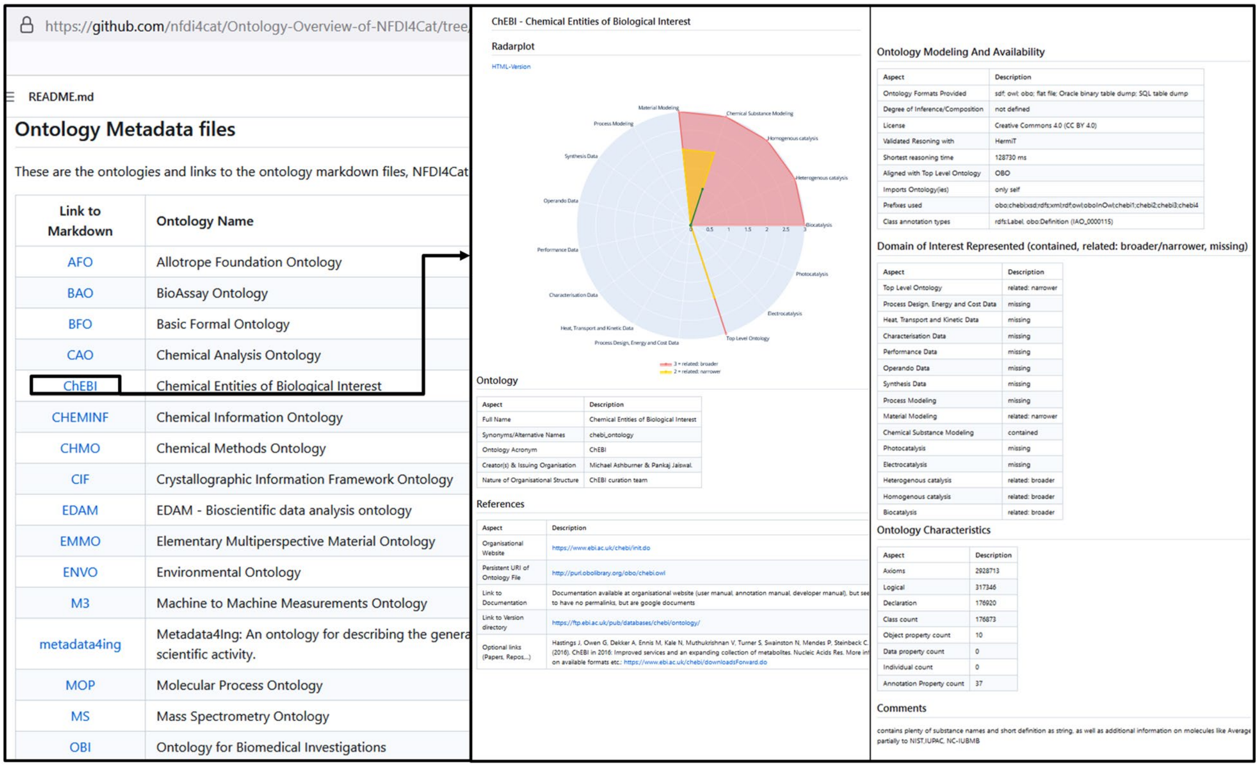


Fig. 1 Radar plot for the amount of ontologies that address the respective domains of catalysis research. The specific fields of catalysis are denoted in blue, while the fields more directed to modelling are colored purple. Fields regarding general catalytic data are written in black

Mapping of ontology classes

The mapping of classes between ontology pairs is performed for the investigated ontologies except for OntoCAPE. This is due to the modular and deprecated design of the ontology, which made it impossible for the authors to load it properly with owlready2. Thus, a total of 22 ontologies are subject to the mapping described in Sect. "Ontology mappings". Table 3 lists the resulting number of classes mapped for each pair of ontologies. The main diagonal of the table lists the total amount of classes for each ontology. Exemplarly, the method found 107 similar classes (entry underlined in Table 3) in the Allotrope Foundation Ontology [19] (AFO, total of 2876 classes) and BioAssay Ontology [24] (BAO, total of 7512 classes) ontologies. The table of mapped classes is also included in the main *readme* file of the repository, providing an interactive interface via Markdown syntax.

There, numbers within the table are clickable, redirecting users to dedicated Markdown files. These files in turn list the detailed information on the mappings between respective ontology pairs in a table. The first two columns record the class IRI and rationale for the mapping (IRI or associated class attributes) from the first ontology. Similarly, the next two columns document the same data taken from the mapped class of the second ontology. Finally, the last column contains the textual definition of the class of the second ontology, where available. This ensures a comprehensive and transparent documentation of the mapping process, allowing for easy access and review of mapped classes. Figure 3 shows an exemplary excerpt of such a Markdown file containing details of the mapping between the two ontologies AFO [19] and BAO [24]. The overall number of mapped classes between those two ontologies is 107 as listed in Table 3. For simplicity, only five classes are presented here, showing the structure of the resulting mapping tables. First, classes are listed that are mapped based on the same IRI. These classes have the same IRI and thus are mapped accordingly. In this example, from entry 16 on, the mappings based on the associated class attributes are listed. The class with the rdfs:label *Shape* in the AFO is mapped to a class in the BAO with a skos:altLabel entry *Shape*. Furthermore, the next listed class with the skos:altLabel *time* in the AFO is mapped to a class in the BAO with a



https://github.com/nfdi4cat/Ontology-Overview-of-NFDI4Cat/tree

README.md

Ontology Metadata files

These are the ontologies and links to the ontology markdown files, NFDI4Cat

Link to Markdown	Ontology Name
AFO	Allotrope Foundation Ontology
BAO	BioAssay Ontology
BFO	Basic Formal Ontology
CAO	Chemical Analysis Ontology
ChEBI	Chemical Entities of Biological Interest
CHEMINF	Chemical Information Ontology
CHMO	Chemical Methods Ontology
CIF	Crystallographic Information Framework Ontology
EDAM	EDAM - Bioscientific data analysis ontology
EMMO	Elementary Multiperspective Material Ontology
ENVO	Environmental Ontology
M3	Machine to Machine Measurements Ontology
metadata4ing	Metadata4Ing: An ontology for describing the genera scientific activity.
MOP	Molecular Process Ontology
MS	Mass Spectrometry Ontology
OBI	Ontology for Biomedical Investigations

ChEBI - Chemical Entities of Biological Interest

Radarplot

HTML-Version

Ontology

Aspect	Description
Full Name	Chemical Entities of Biological Interest
Synonyms/Alternative Names	chebi_ontology
Ontology Acronym	ChEBI
Creator(s) & Issuing Organisation	Michael Ashburner & Pankaj Jaiswal.
Nature of Organisational Structure	ChEBI curation team

References

Aspect	Description
Organisational Website	https://www.ebi.ac.uk/chebi/init.do
Persistent URI of Ontology File	http://purl.obolibrary.org/obo/chebi.owl
Link to Documentation	Documentation available at organisational website (user manual, annotation manual, developer manual), but see to have no permalinks, but are google documents
Link to Version directory	https://ftp.ebi.ac.uk/pub/databases/chebi/ontology/
Optional links (Papers, Repos,...)	Hastings J, Owen G, Dekker A, Ennis M, Kale N, Muthukrishnan V, Turner S, Swainston N, Mendes P, Steinbeck C. (2016). ChEBI in 2016: Improved services and an expanding collection of metabolites. Nucleic Acids Res. More in on available formats etc.: https://www.ebi.ac.uk/chebi/downloadsForward.do

Ontology Modeling And Availability

Aspect	Description
Ontology Formats Provided	sdf; owl; obo; flat file; Oracle binary table dump; SQL table dump
Degree of Inference/Composition	not defined
License	Creative Commons 4.0 (CC BY 4.0)
Validated Resoning with	HermiT
Shortest reasoning time	128730 ms
Aligned with Top Level Ontology	OBO
Imports Ontology(ies)	only self
Prefixes used	obo;chebi;xsd;rdfs;xml;rdf;owl;oboInOwl;chebi1;chebi2;chebi3;chebi4
Class annotation types	rdfs:Label, obo:Definition (IAO_0000115)

Domain of Interest Represented (contained, related: broader/narrower, missing)

Aspect	Description
Top Level Ontology	related: narrower
Process Design, Energy and Cost Data	missing
Heat, Transport and Kinetic Data	missing
Characterisation Data	missing
Performance Data	missing
Operando Data	missing
Synthesis Data	missing
Process Modeling	missing
Material Modeling	related: narrower
Chemical Substance Modeling	contained
Photocatalysis	missing
Electrocatalysis	missing
Heterogenous catalysis	related: broader
Homogenous catalysis	related: broader
Biocatalysis	related: broader

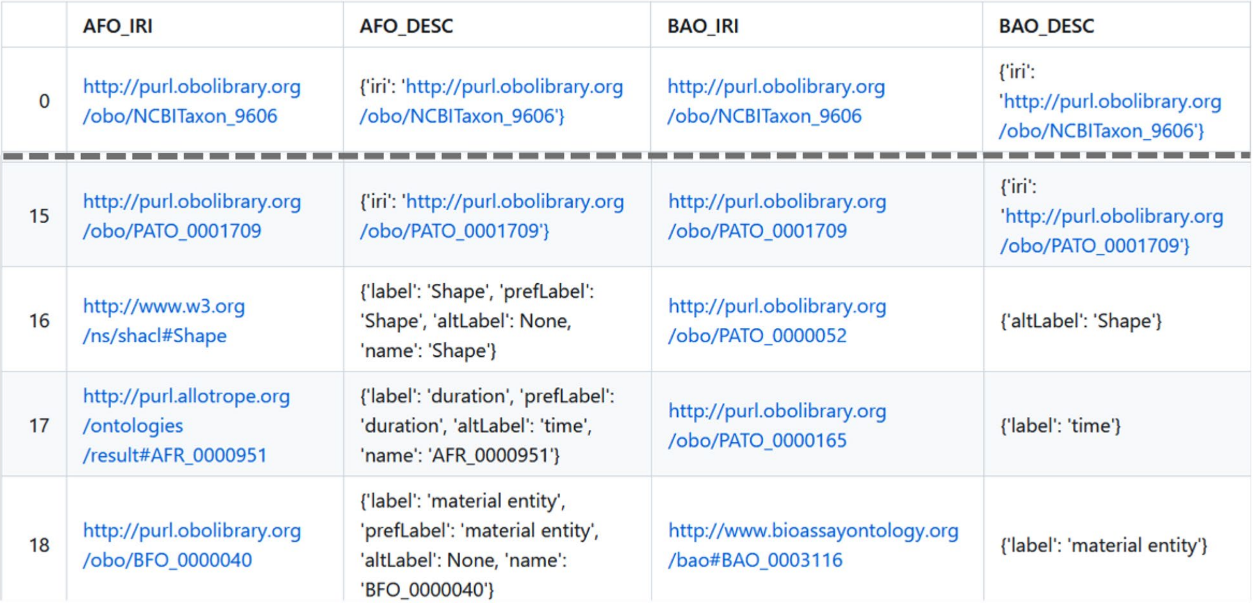
Ontology Characteristics

Aspect	Description
Axioms	2928713
Logical	317346
Declaration	176920
Class count	176873
Object property count	10
Data property count	0
Individual count	0
Annotation Property count	37

Comments

contains plenty of substance names and short definition as string, as well as additional information on molecules like Average partially to NIST,IUPAC, NC-IUBMB

Fig. 2 Visualization of the ontology classification via Markdown files on GitHub. The Markdown rendering of the repository *readme* file (left) lists the ontologies and links to the Markdown files describing the respective ontology (right) according to the classifications listed in Table 1 as well as the radar plot, visualizing the respective domains of catalysis research specific to the ontology (top right)



	AFO_IRI	AFO_DESC	BAO_IRI	BAO_DESC
0	http://purl.obolibrary.org/obo/NCBITaxon_9606	{'iri': 'http://purl.obolibrary.org/obo/NCBITaxon_9606'}	http://purl.obolibrary.org/obo/NCBITaxon_9606	{'iri': 'http://purl.obolibrary.org/obo/NCBITaxon_9606'}
15	http://purl.obolibrary.org/obo/PATO_0001709	{'iri': 'http://purl.obolibrary.org/obo/PATO_0001709'}	http://purl.obolibrary.org/obo/PATO_0001709	{'iri': 'http://purl.obolibrary.org/obo/PATO_0001709'}
16	http://www.w3.org/ns/shacl#Shape	{'label': 'Shape', 'prefLabel': 'Shape', 'altLabel': None, 'name': 'Shape'}	http://purl.obolibrary.org/obo/PATO_0000052	{'altLabel': 'Shape'}
17	http://purl.allotrope.org/ontologies/result#AFR_0000951	{'label': 'duration', 'prefLabel': 'duration', 'altLabel': 'time', 'name': 'AFR_0000951'}	http://purl.obolibrary.org/obo/PATO_0000165	{'label': 'time'}
18	http://purl.obolibrary.org/obo/BFO_0000040	{'label': 'material entity', 'prefLabel': 'material entity', 'altLabel': None, 'name': 'BFO_0000040'}	http://www.bioassayontology.org/bao#BAO_0003116	{'label': 'material entity'}

Fig. 3 Exemplary excerpt of mapping of AFO and BAO ontologies, converted as Markdown file and rendered via GitHub. The grey dashed line denotes a jump in the list, as the first 16 entries (entry 0–15) show mappings because of same class IRIs, while the following entries show mappings due to same annotations of classes

Table 3 Resulting number of classes mapped for each pair of ontologies

	AFO	BAO	BFO	CAO	ChEBI	CHEMINF	CHMO	CIF	EDAM	EMMO	ENVO	M3	M4I	MOP	MS	OBI	OM	OSMO	REX	RXNO	SBO	VIMMP
AFO	2876																					
BAO	107	7512																				
BFO	36	4	35																			
CAO	121	25	14	445																		
ChEBI	58	1678	1	45	176873																	
CHEMINF	92	14	35	42	2	850																
CHMO	249	39	12	69	23	19	3101															
CIF	128	24	4	19	12	37	6	32														
EDAM	50	35	0	12	3	19	9	15	3473													
EMMO	144	21	4	22	23	36	10	502	14	935												
ENVO	248	212	26	84	939	63	36	62	21	64	6566											
M3	88	27	0	19	9	8	6	67	2	65	389	761										
M4I	18	2	3	7	1	4	3	7	1	13	5	10	32									
MOP	6	7	3	8	58	3	3	3	0	3	25	0	1	3686								
MS	140	47	0	26	20	28	30	35	26	31	75	61	1	1	14989							
OBI	289	172	35	82	136	236	77	61	48	54	399	97	6	6	55	4866						
OM	100	21	1	17	11	21	2	80	5	78	226	131	4	0	24	81	815					
OSMO	8	1	0	2	0	8	0	5	4	4	13	4	1	0	3	19	5	173				
REX	9	7	0	2	0	0	18	0	0	1	16	6	1	23	2	7	1	0	552			
RXNO	14	8	2	17	230	5	10	5	0	4	95	1	1	123	3	16	0	0	12	1019		
SBO	41	27	2	7	13	9	3	14	7	17	62	11	1	20	9	31	31	1	11	8	694	
VIMMP	83	13	3	19	3	33	5	83	15	90	74	37	8	1	12	96	72	172	0	2	9	1082

The main diagonal of the table lists the total number of classes contained in the respective ontology. Exemplary for the pair of AFO and BAO ontologies, 107 similar classes were found (underlined in table). An interactive version of this table via Markdown can be found in the main *readme* of the GitHub repository [52] rdfs:label entry *time*. By clicking on the IRI, users are can get deeper insights on the ontology class as hosted by the ontology providers.

Conclusion

This work presents a workflow to setup metadata of ontologies with focus on the domain relevance and display current data for comparison. The metadata is recorded with regards to specific domains of knowledge that extends the data usually presented in ontology databases such as EBI OLS [6] and BioPortal [7]. Furthermore, a codebase is presented that transfers the collected metadata automatically into easy to read Markdown files. Integration into GitHub facilitates visual representation of the metadata and provides quick insight into those ontologies that are most relevant to a particular knowledge domain. The metadata and comparison is made accessible through a GitHub repository and also exported as JSON files for machine-readability. The overall systematic method offers efficient means of comparison across ontologies from a domain of knowledge. Thus, the implemented code and metadata templates aim to be as reusable as possible, to allow for further adaptation on other domains of knowledge.

By dividing into 14 subdomains in three areas, this way of collecting ontology metadata is shown for the domain of catalysis research. With this, a total of 30 ontologies were selected for further screening, but seven were excluded due to accessibility issues or lack of relevance to the domain. The remaining 23 were investigated revealing varying levels of complexity and coverage across different domains within catalysis research. The classification included, among others, the relatedness of the ontologies to each of the 14 subdomains. Relatedness to each subdomain was ranked by four categories; contained, related:narrower, related:broader, and missing. This revealed a graphic representation of the ontologies' metadata for catalysis research, as depicted in Fig. 1. While four subdomains were connected to multiple ontologies, ten were only modeled by one or none. Furthermore, some ontologies posed challenges in reasoning and have differing levels of expressivity. This emphasizes the need for more ontologies or more extended ontologies to describe the domain of catalysis research in more detail.

An approach for automated mapping of classes between ontologies is described, showing potential mappings between classes of overall 22 ontologies related to catalysis research. The results of the mapping are also represented in GitHub for better accessibility and readability in Markdown files. Moreover, Markdown files are created for each pair of ontologies, listing the classes and reasons for mapping of the classes of both mapped ontologies for further review.

While searching for similar class annotations might give a hint on possible class mappings between two ontologies, a user-controlled revision of these mappings should take place. As this task is quite tedious, automation of this process with other code-based solutions should be investigated further. For example, a comparison of the textual, often sentence-wise definitions of a class could be taken into account. A promising technique is described by Korel et al. [53] which could be used in future work to help in automated mappings of ontology classes by similar textual definitions. However, this will only help in mappings of classes, where those definitions are provided, which is often not the case. Here, mapping techniques could be applied, that also considers the interconnection of the class candidates in their respective ontologies.

Acknowledgements

The authors thank the Deutsche Forschungsgemeinschaft (DFG) for funding this research as part of the Nationale Forschungsdateninfrastruktur (NFDI) initiative (grant No.: NFDI/2-1-2021). Furthermore, the authors thank the members of TA1 from NFDI4Cat, especially David Linke and Mark Doerr for their thoughtful insights with regards to this work are acknowledged. The authors would also like to thank Philip Strömert for his helpful comments on the metadata collection and especially early pushed issues on GitHub. A.S.B. thanks the networking program 'Sustainable Chemical Synthesis 2.0' (SusChemSys 2.0) for the support and fruitful discussions across disciplines.

Author contributions

ASB: conceptualization, data curation, methodology, software, validation, investigation, data curation, writing—original draft, writing—review and editing, visualization. HB: conceptualization, data curation, methodology, writing—review and editing. NK: conceptualization, funding acquisition, supervision, writing—review and editing. All authors read and approved the final manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. The authors thank the Deutsche Forschungsgemeinschaft (DFG) for funding this research as part of the Nationale Forschungsdateninfrastruktur (NFDI) initiative (grant No.: NFDI/2-1-2021) for catalysis Research (NFDI4Cat).

Availability of data and materials

The files contained in the following GitHub directory are available free of charge. The ontology metadata taken up for this work and the code developed and described in this work are available in a GitHub repository at: https://github.com/nfdi4cat/Ontology-Overview-of-NFDI4Cat The metadata collection might be subject to change due to further contributions. Thus, the state of the repository as described in this work is available on Zenodo https://zenodo.org/doi/10.5281/zenodo.10470987 convert.py: contains the Python methods described in Sect. "Documentation of the recorded metadata" to convert the ontology metadata to the respective Markdown files and generate the radar plots. similarities.py: contains the Python methods described in Sect. "Ontology mappings" to conduct the automated mapping of classes of different ontologies. The repositories subdirectory ./master_table/ contains the Microsoft Excel file to take up the ontology metadata.

Declarations

A part of this work with preliminary and deprecated results has been published as conference talk and its abstract [54].

Competing interests

The authors declare that they have no competing interests.

Received: 25 October 2023 Accepted: 22 January 2024
Published online: 07 February 2024

References

1. Wilkinson MD et al (2016) The FAIR guiding principles for scientific data management and stewardship. Sci Data 3:160018. https://doi.org/10.1038/sdata.2016.18 **[cito:citesAsAuthority] [cito:agreesWith]**
2. Gruber TR (1993) A translation approach to portable ontology specifications. Knowl Acquis 5:199–220. https://doi.org/10.1006/knac.1993.1008 **[cito:citesAsAuthority] [cito:agreesWith]**
3. Wulf C et al (2021) A unified research data infrastructure for catalysis research–challenges and concepts. ChemCatChem 13:3223–3236. https://doi.org/10.1002/cctc.202001974 **[cito:citesAsAuthority] [cito:agreesWith]**
4. Trunschke A (2022) Prospects and challenges for autonomous catalyst discovery viewed from an experimental perspective. Catal Sci Technol 12:3650–3669 **[cito:citesAsAuthority] [cito:agreesWith]**
5. Horsch M et al (2022) Interoperability and architecture requirements analysis and metadata standardization for a research data infrastructure in catalysis. In: Pozanenko A, Stupnikov S, Thalheim B, Mendez E, Kiselyova N (eds) Data analytics and management in data intensive domains, Vol. 1620 of communications in computer and information science, Springer International Publishing, Cham, p 166–177 . https://doi.org/10.1007/978-3-031-12285-9_10 **[cito:usesDataFrom] [cito:extends]**
6. Jupp S, Burdett T, Leroy C, Parkinson HE, Malone J, Stevens R, Forsberg K, Splendiani A (2015) A new ontology lookup service at EMBL-EBI. In: Malone J, Stevens R, Forsberg K, Splendiani A (eds) Proceedings of the 8th Semantic Web Applications and Tools for Life Sciences International Conference, Cambridge UK, December 7-10, 2015, Vol. 1546 of CEUR Workshop Proceedings, p 118–119. https://ceur-ws.org/Vol-1546/paper_29.pdf **[cito:citesAsDataSource] [cito:usesDataFrom]**. Accessed 12 Dec 2023
7. Noy NF et al (2009) BioPortal: ontologies and integrated data resources at the click of a mouse. Nucleic Acids Res 37:W170-3. https://doi.org/10.1007/978-3-031-12285-9_10 **[cito:citesAsDataSource] [cito:usesDataFrom]**
8. Strömert P, Hunold J, Castro A, Neumann S, Koepler O (2022) Ontologies4Chem: the landscape of ontologies in chemistry. Pure Appl Chem 94:605–622. https://doi.org/10.1515/pac-2021-2007 **[cito:agreesWith] [cito:extends] [cito:usesDataFrom] [cito:citesAsDataSource]**
9. Alliance for Internet of Things Innovation (2021) Ontology landscape. https://aioti.eu/wp-content/uploads/2022/02/AIOTI-Ontology-Landscape-Report-R1-Published-1.0.1.pdf **[cito:citesAsDataSource]**. Accessed 12 Dec 2023
10. OBOFoundry. OBO Dashboard. http://dashboard.obofoundry.org/dashboard/index.html. Accessed 10 Oct 2023. **[cito:discusses] [cito:agreesWith]**
11. Jackson R et al (2021) OBO foundry in 2021: operationalizing open data principles to evaluate ontologies. Database. https://doi.org/10.1093/database/baab069 **[cito:citesAsAuthority]**
12. Prud'hommeaux E, Carothers G (2014) RDF 1.1 Turtle. W3C Recommendation, W3C. https://www.w3.org/TR/2014/REC-turtle-20140225/ **[cito:citesAsAuthority]**. Accessed 12 Dec 2023
13. Krötzsch M, Patel-Schneider P, Hitzler P, Parsia B, Rudolph S (2012) OWL 2 web ontology language primer (2nd edn). W3C Recommendation, W3C. https://www.w3.org/TR/2012/REC-owl2-primer-20121211/ **[cito:citesAsAuthority]**. Accessed 12 Dec 2023
14. Arp R, Smith B, Spear AD (2015) Building ontologies with Basic Formal Ontology (Massachusetts Institute of Technology, Cambridge, Massachusetts, 2015) **[cito:discusses] [cito:citesAsAuthority]**
15. Musen MA (2014) The Protégé project: a look back and a look forward. AI Matters 1:4–12 **[cito:usesMethodIn]**
16. McKinney W et al (2010) Data structures for statistical computing in Python **[cito:usesMethodIn]**
17. Lamy J-B (2017) Owlready: ontology-oriented programming in Python with automatic classification and high level constructs for biomedical ontologies. Artif Intell Med 80:11–28. https://doi.org/10.1093/database/baab069 **[cito:usesMethodIn]**
18. Jackson RC et al (2019) ROBOT: a tool for automating ontology workflows. BMC Bioinform 20:407. https://doi.org/10.1186/s12859-019-3002-3 **[cito:usesMethodIn]**
19. Allotrope Foundation (2018) Allotrope Foundation Ontologies. https://www.allotrope.org/ontologies **[cito:discusses]**. Accessed 12 Dec 2023
20. Batchelor C (2012) Chemical analysis ontology. https://github.com/rsc-ontologies/rsc-cmo **[cito:discusses]**. Accessed 12 Dec 2023
21. Leal D (2005) ISO 15926 life cycle data for process plant: an overview. Oil Gas Sci Technol 60:629–637. https://doi.org/10.2516/ogst:2005045 **[cito:discusses]**
22. Bandrowski A et al (2016) The ontology for biomedical investigations. PLoS ONE 11:e0154556. https://doi.org/10.1371/journal.pone.0154556 **[cito:discusses]**
23. Horsch MT et al (2021) OSMO: ontology for simulation, modelling, and optimization. https://doi.org/10.5281/zenodo.5084393 **[cito:discusses]**
24. Visser U et al (2011) BioAssay ontology (BAO): a semantic description of bioassays and high-throughput screening results. BMC Bioinform 12:257. https://doi.org/10.1186/1471-2105-12-257 **[cito:discusses]**
25. Friis J et al (2023) emmo-repo/cif-ontology: v0.1.0. https://doi.org/10.5281/zenodo.7966648**[cito:discusses]**
26. ISO 15926-14: 2020 (2020) Industrial automation systems and integration–integration of life-cycle data for process plants including oil and gas production facilities—part 14: industrial top level ontology. https://www.iso.org/standard/75949.html **[cito:discusses]**. Accessed 12 Dec 2023
27. Fumagalli L, Pala S, Garetti M, Negri E (2014) Ontology-based modeling of manufacturing and logistics systems for a new MES architecture. 8827:192–200. https://doi.org/10.1007/978-3-662-44739-0_24 **[cito:discusses]**
28. Horsch MT (2023) PIMS-II ontology. Version II.1.12a. http://www.molmod.info/semantics/pims-ii/ **[cito:discusses]**. Accessed 12 Dec 2023
29. Horsch MT, Schembera B (eds) (2022) Documentation of epistemic metadata by a mid-level ontology of cognitive processes: Zenodo. https://doi.org/10.5281/zenodo.6638457 **[cito:discusses]**
30. Borgo S et al (2022) DOLCE: A descriptive ontology for linguistic and cognitive engineering1. Appl Ontol 17:45–69. https://doi.org/10.3233/AO-210259 **[cito:discusses]**
31. Gyrard A, Datta SK, Bonnet C, Boudaoud K (2015) Cross-domain internet of things application development: M3 framework and evaluation 9–16. https://doi.org/10.1109/FiCloud.2015.10 **[cito:discusses]**
32. Rijgersberg H, Wigham M, Top J (2011) How semantics can improve engineering processes: a case of units of measure and quantities. Adv Eng Inform 25:276–287. https://doi.org/10.1016/j.aei.2010.07.008 **[cito:discusses]**
33. Degtyarenko K (2007) REX ontology of physico-chemical processes. http://purl.obolibrary.org/obo/rex.owl **[cito:discusses]**. Accessed 12 Dec 2023
34. Chalk S, Williams A (2015) Chemical analysis ontology. https://champ.stuchalk.domains.unf.edu/cao **[cito:discusses]**. Accessed 12 Dec 2023
35. Black M et al (2022) EDAM: the bioscientific data analysis ontology. https://doi.org/10.7490/F1000RESEARCH.1118900.1 **[cito:discusses]**
36. Arndt S et al (2023) Metadata4ing: an ontology for describing the generation of research data within a scientific activity. https://doi.org/10.5281/zenodo.5957103 **[cito:discusses]**
37. Marquardt W (2010) OntoCAPE: a re-usable ontology for chemical process engineering RWTH edition, Springer, Heidelberg **[cito:discusses]**
38. Batchelor C (2012) Chemical reactions ontology (RXNO). https://github.com/rsc-ontologies/rxno **[cito:discusses]**. Accessed 12 Dec 2023
39. Hastings J, Owen G, Dekker A, Ennis M, Kale N, Muthukrishnan V, Turner S, Swainston N, Mendes P, Steinbeck C (2016) ChEBI in 2016: improved services and an expanding collection of metabolites. Nucleic Acids Res 44:D1214–D1219. https://doi.org/10.1093/nar/gkv1031 **[cito:discusses]**
40. Hashibon A, Ghedini E, Schmitz G, Goldbeck G, Friis J (2022) Elemental multiperspective material ontology. http://emmo.info/emmo **[cito:discusses]**. Accessed 12 Dec 2023
41. Batchelor C (2012) Molecular process ontology (MOP). https://github.com/rsc-ontologies/rxno **[cito:discusses]**. Accessed 12 Dec 2023

42. Krdzavac N et al (2019) An ontology and semantic web service for quantum chemistry calculations. J Chem Inform Model 59:3154–3165. https://doi.org/10.1021/acs.jcim.9b00227 **[cito:discusses]**
43. Juty N, Le Novère N (2013) Systems biology ontology. In: Dubitzky W, Wolkenhauer O, Cho K-H, Yokota H (eds) Encyclopedia of systems biology 2063, Springer Reference, New York. https://doi.org/10.1007/978-1-4419-9863-7_1287 **[cito:discusses]**
44. Hastings J et al (2011) The chemical information ontology: provenance and disambiguation for chemical data on the biological semantic web. PLoS ONE 6:e25513. https://doi.org/10.1371/journal.pone.0025513 **[cito:discusses]**
45. Buttigieg PL, Morrison N, Smith B, Mungall CJ, Lewis SE (2013) The environment ontology: contextualising biological and biomedical entities. J Biomed Semant 4:43. https://doi.org/10.1186/2041-1480-4-43 **[cito:discusses]**
46. Mayer G, Montecchi-Palazzi L, Ovelleiro D, Jones AR, Binz PA, Deutsch EW, Chambers M, Kallhardt M, Levander F, Shofstahl J, Orchard S (2013) The HUPO proteomics standards initiative- mass spectrometry controlled vocabulary. Database 2013:bat009. https://doi.org/10.1093/database/bat009 **[cito:discusses]**
47. Farazi F et al (2020) OntoKin: an ontology for chemical kinetic reaction mechanisms. J Chem Inf Model 60:108–120. https://doi.org/10.1021/acs.jcim.9b00960 **[cito:discusses]**
48. Horsch MT et al (2021) Introduction to the VIMMP ontologies. https://doi.org/10.5281/zenodo.3936795 **[cito:discusses]**
49. Plotly Technologies Inc. (2015) Collaborative data science. https://plot.ly **[cito:usesMethodIn]**. Accessed 12 Dec 2023
50. Glimm B, Horrocks I, Motik B, Stoilos G, Wang Z (2014) HermiT: an OWL 2 reasoner. J Autom Reason 53:245–269. https://doi.org/10.1007/s10817-014-9305-1 **[cito:usesMethodIn]**
51. Tsarkov D, Horrocks I, Furbach U, Shankar N (2006) FaCT++ description logic reasoner: system description. In: Furbach U, Shankar N (eds) Lecture Notes in Computer Science, Vol. 4130, Springer-Verlag GmbH, Berlin Heidelberg, p 292–297. https://doi.org/10.1007/11814771_26 **[cito:usesMethodIn]**
52. Behr AS, Borgelt H (2023) Github: ontology overview of NFDI4Cat. https://github.com/nfdi4cat/Ontology-Overview-of-NFDI4Cat **[cito:discusses]**. Accessed 12 Dec 2023
53. Korel L, Yorsh U, Behr AS, Kockmann N, Holeňa M, (2013) Text-to-ontology mapping via natural language processing with application to search for relevant ontologies in catalysis. Computers. https://doi.org/10.3390/computers12010014 **[cito:discusses] [cito:agreesWith]**
54. Behr AS, Borgelt H, Petrenko T, Dörr M, Kockmann N (2023) Investigating the landscape of ontologies for catalysis research data management. In: Proceedings of the Conference on Research Data Infrastructure 1. https://doi.org/10.52825/cordi.v1i.232 **[cito:extends]**

Publisher's Note