Development of an Ontology for Biocatalysis

Marian J. Menke[1,‡], Alexander S. Behr[2,‡], Katrin Rosenthal[3], David Linke[4], Norbert Kockmann[2], Uwe T. Bornscheuer[1,*], and Mark Dörr[1,*]

DOI: 10.1002/cite.202200066

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

Enzyme activity data for biocatalytic applications are currently often not annotated with standardized conditions and terms. This makes it extremely hard to retrieve, compare, and reuse enzymatic data. With advances in the fields of artificial intelligence (AI) and machine learning (ML), the automated usability of data in the form of machine-readable annotations will play a crucial role for their success. It is becoming increasingly easy to retrieve complex data sets and extract relevant information; however, standardized data readability is a current limitation. In this contribution, we outline an iterative approach to develop standardized terms and create semantic relations (ontologies) to achieve this highly desirable goal of improving the discoverability, accessibility, interoperability, and reuse of digital resources in the field of biocatalysis.

Keywords: Biocatalysis, FAIR data, Metadata standard, Ontology, Semantics

Received: May 16, 2022; *revised:* July 22, 2022; *accepted:* August 18, 2022

1 Introduction

Modern biocatalysis includes a wide variety of different scientific fields, such as biology, chemistry, engineering, and recently more and more often computer and data science, in order to achieve a much deeper understanding of the relationships between protein sequence, structure, activity, and reaction mechanism [1]. Furthermore, the enormous product variety in numerous industrial branches and increasing demand of greener catalysts for the fabrication of high-value products require a computational-based systematic approach to protein engineering [2, 3]. The necessity for interdisciplinarity of these scientific fields makes it intellectually attractive and can be a gift [4], but the resulting complexity can also be a burden. Currently, interoperability of information about different scientific fields is generally hard to achieve, and especially in the wide area of biocatalytic data, results of different laboratories are hardly comparable due to several reasons.

Firstly, reaction parameters are often only reported incompletely – if they are given at all (which is frequently true for industrially supported research). Halling et al. recently investigated the reproducibility of experiments. They concluded that a lot of published articles are lacking information or only give incomplete information on the experimental setup that would be necessary for reproduction of the results [5]. However, biocatalytic data values are always tightly linked to the experimental setup, such as temperature, stirring speed, reactants and buffer concentrations or oxygen saturation to just name a few. Hence, to reproduce a biocatalytic experiment, it is necessary to have a complete list of reaction conditions and devices. Incidentally, the problem of missing information to generate data and thus reproducibility is not unique to biocatalysis but can be found in any field of study [6]. In addition, many of the reported data do not contain information on how they were obtained and how reliable they are [7].

Secondly, reaction conditions are often chosen by the experience of a certain scientist or are to some extend adjusted to optimize the turnover for a particular reaction, which leads to many fine-grained differences of experimental protocols, which directly lead to hardly comparable data sets. There exist no standards for reaction conditions to achieve comparability of, e.g., enzyme activities of related enzymes, like benchmark, but also scientists often do not accept reporting standards, like STRENDA [8] and even if they do, the published data is almost impossible to be processed

[1]Marian J. Menke, Prof. Dr. Uwe T. Bornscheuer, Dr. Mark Dörr
uwe.bornscheuer@uni-greifswald.de,
mark.doerr@uni-greifswald.de
University of Greifswald, Dept. of Biotechnology & Enzyme Catalysis, Felix-Hausdorff-Strasse 4, 17487 Greifswald, Germany.

[2]Alexander S. Behr, Prof. Dr. Norbert Kockmann
TU Dortmund University, Department of Biochemical and Chemical Engineering, Laboratory of Equipment Design, Emil-Figge-Strasse 68, 44227 Dortmund, Germany.

[3]Dr. Katrin Rosenthal
TU Dortmund University, Department of Biochemical and Chemical Engineering, Chair for Bioprocess Engineering, Emil-Figge-Strasse 66, 44227 Dortmund, Germany.

[4]Dr. David Linke
Leibniz-Institut für Katalyse e. V., Albert-Einstein-Strasse 29A, 18059 Rostock, Germany.

[‡]These authors contributed equally.

 © 2022 The Authors. Chemie Ingenieur Technik published by Wiley-VCH GmbH

systematically by computer algorithms, since it is not published in formats, algorithms can easily "understand".

In a recent study, Calvelage et al. collected reaction data for almost 500 amine transaminase catalyzed reactions through literature search in order to systemically gather information on the conversion scope of different transaminases and group them into different classes [9]. It quickly became apparent that not only the collection of this data is tedious because descriptive information is usually given in a plain text form buried only within the human readable article, but more importantly the gathered data could not be compared because of the different experimental conditions. Since the description of data can additionally be separated into tables, figures, and diagrams, it is (almost) impossible to rely on an algorithm to extract all the necessary information [10]. The existence of a database for biocatalytic data, similar to the protein database (pdb.org), would speed up the often time-consuming and tedious literature screening process a researcher has to perform to support the research activities. An additional side-effect of a database would also be the normalization of data, which would overall yield more uniform data, as researchers can re-utilize the previously obtained data. To define the term, in this article 'biocatalysis' is mainly understood as the application of isolated enzymes in free or immobilized form to convert substrates into products.

1.1 Existing Data Bases and Approaches

Several databases are already available for different purposes, i.e., gene annotations in GenBank, enzymatic information in BRENDA, protein structure and sequence information in Uniprot and crystallographic data in the Protein Data Bank (PDB) [11] with some of them being also interconnected [12]. However, no database exists to date that allows for a quick search of biocatalytic activity data, hence, the literature has to be screened manually. First steps into this direction are done with EnzymeML and its graphical user interface BioCatHub towards standardized biocatalytic data and web-based storage to allow for sharing of scientific data via XML (Extensible Markup Language) documents [10, 13–14]. Accessibility and integration of data is important to not only assure reproducibility [15], but also to enable new findings with already available data sets. The value of reusing data can be exemplified on the recently published machine learning driven biocatalysis synthesis planning pipeline, which revolved around data-driven prediction of retro-synthesis routes [16], as well as on the Galaxy-SynBioCaD pipeline for synthetic biology and metabolic engineering [17].

The growing amount of experimental data, which can already be complicated for a human to evaluate and read through, is becoming a more and more apparent problem [18]. Thus, a digital solution needs to be found and implemented to solve this growing problem. In 2016, the FAIR guidelines were introduced for improved data management [11], which should pave the way to better operability of experimental data. FAIR stands for findability, accessibility, interoperability, and reusability and describes machine-guided data extraction, accession, and integration to other data.

With these FAIR guidelines in mind, the German NFDI initiative (Nationale Forschungsdateninfrastruktur, national research data infrastructure) was founded in October 2020 in order to improve the interdisciplinary data exchange [19]. NFDI operates in different (up to 30) consortia (with 10–20 member institutions each) and its goal is to create a data infrastructure (base) for scientific data in order to make the data reusable for everyone, rather than being stored offline with only limited access. NFDI4Cat (Nationale Forschungsdateninfrastruktur for Catalysis-Related Sciences) is one of the consortia established in the first round and responsible for the development of infrastructure supporting exploration, exchange, and access of data related to the wide field of catalysis and biocatalysis [20].

In order to establish a comprehensive data base, common semantic standards have to be defined first. Data has to be converted from human-readable (literature) into computer-readable form (database). This conversion to structured computer-readable data then can be also transformed to human-readable representation, allowing for better FAIR-ness of data. One obstacle is the different association of terms throughout the different scientific fields and different meanings the same term can have. An elegant and well-established way to define these standardized vocabularies is to define taxonomies with the application of standards used in semantic web technology [21]. Involvement of a computer-guided knowledge extraction tool from literature should help with the creation of such vocabularies as basis of an ontology network for biocatalysis.

1.2 Semantic Data Standards/Semantic Web Technology

Utilizing standardization of data can accelerate the development of new bioprocesses and, therefore, be beneficial for the overall costs of a project as well [22]. Definition and usage of data standards is especially important in automation of processes to enable interoperability of data for unification of data exchange [20]. The term semantic relates to the science of meaning [23], thus providing more meaningful information to humans and machines. Semantics add a relation to conceptual ideas. To describe an underlying concept, it is not sufficient to only define the terms that are involved, but rather also the relation and connection between these terms. Humans can deduce relations between terms easily, but machines need to have a defined and formal semantic relation of concepts in order to "understand" the underlying process. By trying to map the human conception of processes onto machines and making them

 © 2022 The Authors. Chemie Ingenieur Technik published by Wiley-VCH GmbH

machine-understandable, the knowledge quality of the machine-generated output is increased.

To ensure a unified approach for semantic data standards the semantic web technology was invented as an extension to the world wide web and part of the World Wide Web Consortium (W3C) standards [24], enabling definition of machine-readable semantic relations. In order to build a semantic network as an ontology, first the underlying terms have to be defined, i.e., a taxonomy is formed, and second the relations between them have to be described. A widely accepted definition of ontology is "a formal specification of a shared conceptualization" [25], which emphasizes the formal aspect, that means that a formalized, standardized description language should be used to specify terms and their relations (conceptualization), which can be shared between many (humans or machines).

For example, an enzyme can be thought of, or in the above terms "conceptualized" (which is exactly this abstraction process in our thoughts), as a protein-based catalyst, which in turn is a substance, while a biocatalytic reaction is some kind of a catalytic reaction, which in turn can be summarized with other (bio)chemical reactions by a super class called reaction. Every reaction in turn has some reactants, which each can be described as a substance. A catalytic reaction has the specification of at least one catalyst as a necessary reactant. Assuming, that a researcher has conducted a biocatalytic reaction in a certain well of a multiwell plate, this real-world realization of the biocatalytic reaction concept can be described as an individual of the latter class. This short (and incomplete) example of a, yet still not fully formalized, ontology is depicted in Fig. 1.

For a full formalization of this natural language (English) description of the short example ontology above, it needs to be converted into a formal language, which should be human and machine readable. For this purpose the W3C developed, e.g., the Web Ontology Language (OWL) [26, 27].

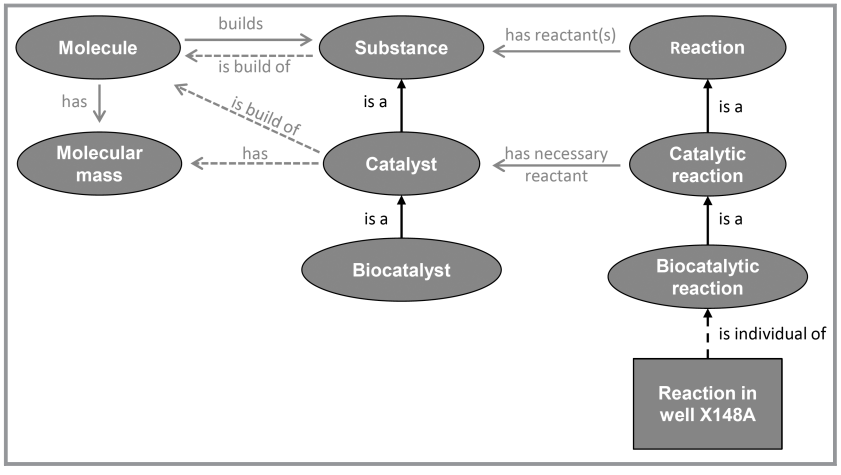

Figure 1. Short example of an ontology. Concepts in ellipsoids denote classes, black arrows hierarchical relations and grey arrows denote relationships between classes, dashed, grey arrows denote inferred relations. An arbitrary reaction conducted, e.g., in a multiwell plate in the laboratory, can then be described as an individual of a biocatalytic reaction, referencing real-world realizations of the concepts defined in classes.

1.3 Ontologies are the Base for a Semantic Network

Many ontologies already exist for specific domains (NFDI4Chem landscape of ontologies in chemistry, Ontologies4Chem [28]; BioPortal, including 1156 ontologies (accessed May 15, 2022) [29]; MatPortal for material science [30]). In order to build a more comprehensive ontology for biocatalysis, it is possible to reuse and combine parts of semantic substructures of existing ontologies. Starting from an already defined, but generally unspecific top-level ontology allows for smooth interoperability and makes the mapping of the custom ontologies easier [31]. Examples of nicely modularized, but large ontologies are the domain ontologies of the Elementary Multiperspective Material Ontology (EMMO) [32] and the BioAssayOntology (BAO) [33] for describing biological and biochemical assays, which is in turn a potential candidate for one "module" of the intended biocatalysis ontology. Currently, concurrent, but unfortunately partially incompatible top-level ontologies exist, e.g., European Materials Modelling Ontology – Top (EMMO-top) [34, 35] or Basic Formal Ontology (BFO) [36]. Choosing a certain top-level ontology constrains the further interoperability of more specific domain ontologies. Therefore, a careful selection is of high importance. After definition of the necessary vocabulary, classes of concepts can be categorized as super-classes, which can be further divided into sub-classes with narrower properties. By doing so, a taxonomic hierarchy of all terms is created. Additionally, definition of properties for each entity of the classes allows for further refinement and assignment of information. A more detailed and comprehensive description of this process can be found in the Handbook on Ontologies [37].

The resulting transitive hierarchical order of classes consists of subclasses with inherited properties of their super classes. This transitive behavior enables the automated generation of relations by inductive logic programming-based inferences leading to improvements, e.g., in information retrieval of data bases [37]. In the simple example above (Fig. 1) the catalyst (and also biocatalyst) property "has molecular mass" can be inferred from the fact that a catalyst is a substance and this is built of a molecule, without explicit definition (mind that in the case of a heterogeneous catalyst, this inference is critical, so ontology design needs to be very careful). In more complex ontologies, very powerful inferences

 © 2022 The Authors. Chemie Ingenieur Technik published by Wiley-VCH GmbH

will be possible, enabling advanced queries of data with associated ontologies, but a lot of care needs to be taken, that these ontologies maintain logically consistent.

Once a custom top-level ontology is created, existing ontologies can be mapped to the self-defined classes, meaning that similarities between the custom and existing ontologies can be found. These mappings are added to the ontology and thus understandable for the computer. This approach allows easy merging of the different ontologies, which then results in a single, large ontology with finer granularity. Next to these taxonomical hierarchy relations, associative links can be defined to show some interconnective dependencies between different classes that are not hierarchically connected, e.g., "is_build_of", "has_reactand" (Fig. 1).

Especially in our case, it is of interest to create a common and uniform taxonomy throughout the different scientific fields that are required to build and characterize a biocatalyst, which is a huge endeavor. For that reason, we started with a simpler ontology approach focusing on the most widely used terms and a few key relations between them using the SKOS (Simple Knowledge Organisation System) standard. In a top-down approach, the main concepts and terms necessary for the overall description of the processes involved in the biocatalysis domain were identified and collected in tables. SKOS enables an easy assignment of different concepts (which are a more generalized form of classes) and values for the entities of each broader (or narrower) concept in a way that allows easy interoperability [38]. The use of SKOS vocabulary as key element of an interoperability framework has been convincingly demonstrated in AgroVoc which is a multi-languange vocabulary for agriculture operated by the United Nations [39]. This SKOS-based approach additionally allows for mapping and merging of already available ontologies, which further refines our created ontology [40].

Since SKOS utilizes a simpler approach compared to OWL ontologies, semantic property relations cannot be rigorously defined besides the definition of concepts that are "broader" and "narrower" to each other. For this reason, in a later stage of ontology development, a more expressive ontology language as OWL will be necessary to extend the vocabulary. OWL is capable of describing relationships of different defined classes to each other. Imagine SKOS annotation of concepts being basic like "concept A is a narrower concept of concept B" (e.g., protein is a concept, with enzyme being a narrower concept of proteins), whereas OWL allows for much more complex descriptive annotations like "concept A is a broader concept of concept B and concept B has a descriptor" (e.g., enzyme is a subclass of protein, build of a molecule with a certain molecular mass. Proteins and enzymes consist of an amino acid sequence, but an amino acid sequence does not have to be a protein, enzyme or biocatalyst.). By specifying logical relations in a formal ontology description language, e.g., OWL, it is possible to logically and automatically deduce many additional properties and relations that are not explicitly defined with a software called reasoner. In our group some experience on deduction of properties and relations from existing ontologies were gained with the reasoner RacerPro [41], FaCT++ [42], and HermiT [43].

1.4 Scope of Ontologies

The power of ontology description languages such as SKOS or OWL is that they are both based on Resource Description Framework (RDF), which allows for linking and connecting of concepts in human- and machine-readable form. Definitions can be assigned as a text to each class to yield a human-readable description. Assignment of different classes and semantics, however, also allow for hierarchical ordering in a machine-readable form, which enables the above explained machine-based logic reasoning.

Deduction and inference are additionally important to help identifying logical inconsistencies in the ontology, so that the user does not end up with contradictory results. Independency of the described entities to the overall ontology enables the addition and growth of the ontology without interfering with previously defined classes and its entities. The overall approach to ontology development from structural and syntactic interoperability to semantic interoperability is visualized in Fig. 2.

2 Development of an Ontology for Biocatalysis

One of the goals, which is addressed by one of the task areas of NFDI4Cat, is to develop an ontology capable of describing the relevant processes important for biocatalysis. A generalized workflow for ontology development for a certain domain is visualized in Fig. 3.

As illustrated in Fig. 4, the ontology development workflow for biocatalysis started with the collection of different processes as concepts, such as cultivation, purification, or enzymatic assays, based on common domain knowledge, textbook knowledge and discussions with domain experts, collected in simple tables. Further definition of different concepts such as the different steps involved for each process refines the taxonomical hierarchy, using SKOS as formal description language. To generate a full ontology, division of these concepts into classes, into subclasses and further refinements of the specific steps of the described process will be formulated in OWL. To each entity of concepts, classes, and subclasses, preferred labels are assigned as the preferred term. Alternative labels as synonyms are collected as well in order to broaden the findability of the exact process later on. Each entity of each subclass can be defined and properties to each entity are assigned. This process should finally result in a knowledge graph that can be used to relate measured data and make data FAIR.

 © 2022 The Authors. Chemie Ingenieur Technik published by Wiley-VCH GmbH

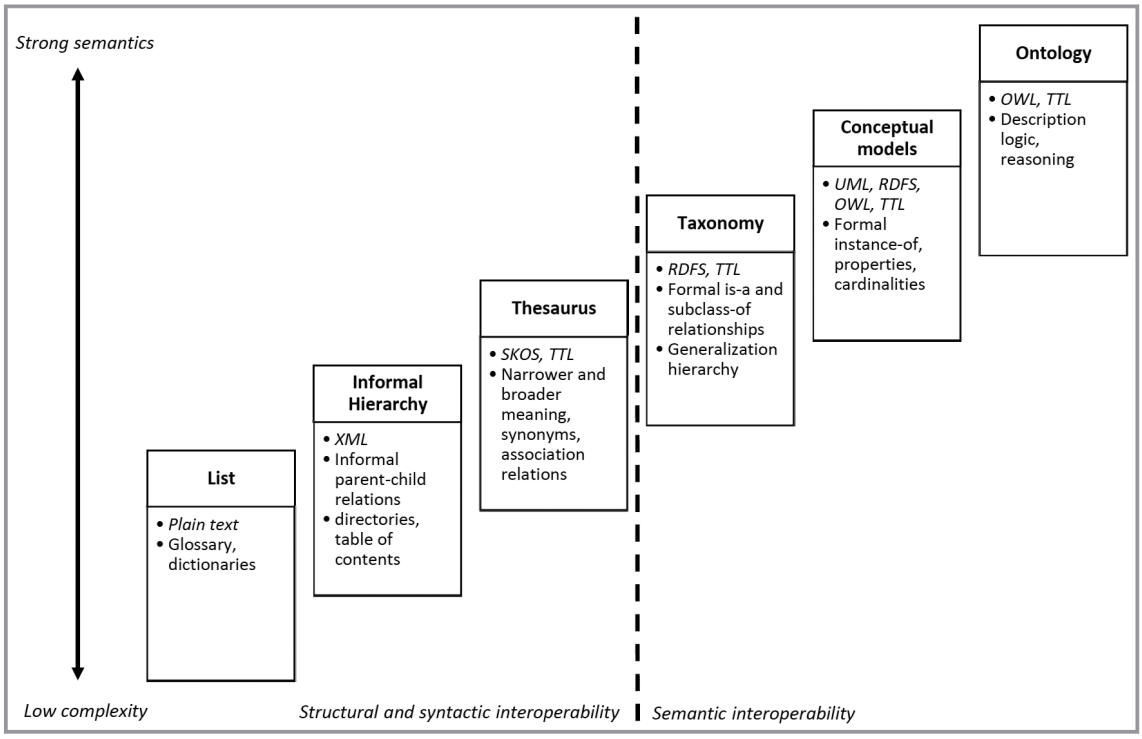

Figure 2. Steps for syntactic and semantic interoperability modified from Maja-Olivia Lenz et al. [44].

Processes relevant to the field of biocatalysis were collected. From this set of entities, definitions are assigned. The concepts get hierarchically sorted by defining skos:broader and skos:narrower relations by assigning subclasses as children of their respective superclasses. All this is conducted in human-readable form, utilizing a spreadsheet template based on VocExcel [45] and NFDI4Cat fork [46]. In addition, entities were named with preferred and alternate labels to take linguistic ambiguities into account, giving unique names as preferred label and providing synonyms if applicable as alternate labels.

A software tool based on python using packages from VocExcel [45], and ontospy [47] is created to read in the concept lists gathered as Excel-files based on a template from VocExcel. It then normalizes the data, creating Uniform Resource Identifier's (URI) for each preferred label listed by the domain experts. This allows also for linkage of classes by the relation of skos:narrower, thus defining the children and creating a hierarchy. The converter module of VocExcel then converts this list to SKOS format and stores it in RDF format, e.g., as turtle file (.ttl, Terse RDF Triple Language). To provide additional human readable documentation, this SKOS-file is read in using the functions HTMLVisualizer and Dataviz of the module ontospy [47]. The output of this call yields two .html documents: an interactive dendrogram (as also shown in Fig. 5) as well as a documentation listing the definitions, close hierarchical structure and RDF/XML implementation of the classes listed in the concept-table. This flow of data is also visualized in Fig. 5. The respective software tool is available on github [48].

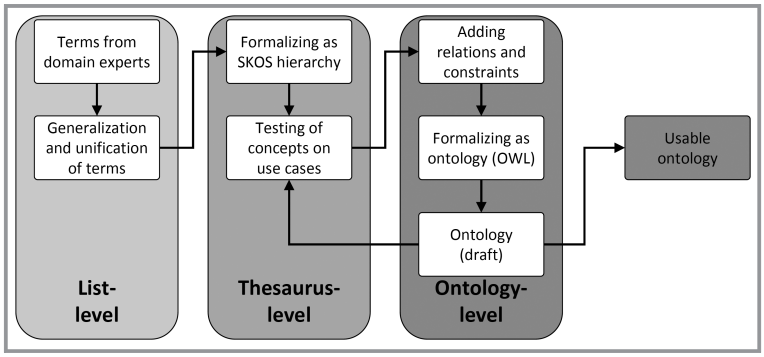

Figure 3. Ontology development design workflow. Compiling a list of necessary terms by domain experts/ enzymologists builds the foundation for further hierarchical ordering of thesauri via SKOS. Checking for completeness and further optimization of the ontology draft is achieved by testing use cases.

Since each entity has also a set of properties, such as conditions and necessary operations needed for the process, we began by defining (generic) classes of conditions and unit operations for the whole biocatalysis ontology, such as speed, time, or temperature. The reasoning behind this is that the entities of the condition and operation classes can be assigned in process_condition and process_unit_operation to a specific process with defined values

 © 2022 The Authors. Chemie Ingenieur Technik published by Wiley-VCH GmbH

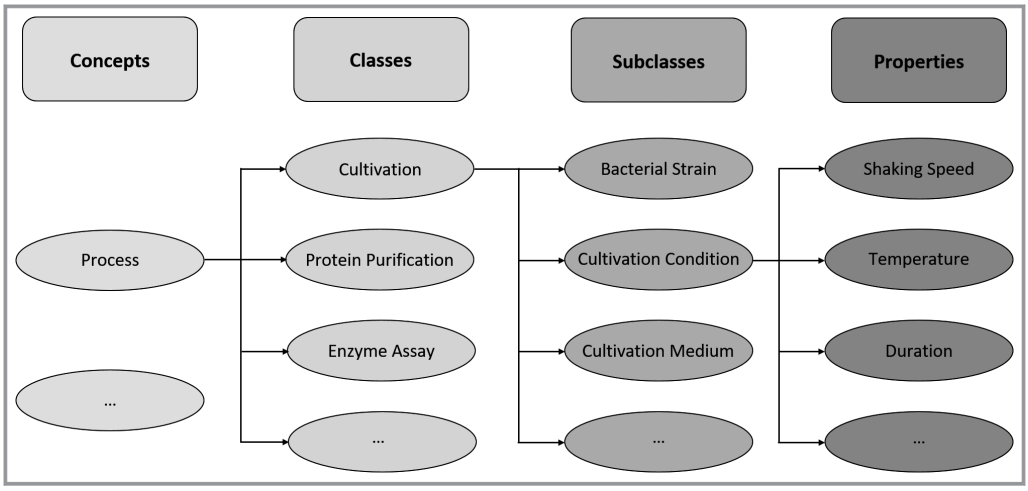

Figure 4. List-level of ontology development for biocatalysis. Terms necessary for the description of processes in biocatalysis are gathered, ordered hierarchically and further refined in granularity by assigning classes, subclasses and properties.

necessary for that process. In that way, we do not need to define all the conditions for each process. The visualization of the hierarchical order of the collected taxonomy as a dendrogram is shown in Fig. 6.

Further semantic relations can then be applied by transforming the SKOS-based hierarchy to an ontology as described in the SKOS-primer [38]. In an iterative process we will then use a standardized ontology manipulation program such as, e.g., Protégé [49], for further iterative manipulation and development of the ontology by adding more semantic relations as well as applying reasoner for consistency checks.

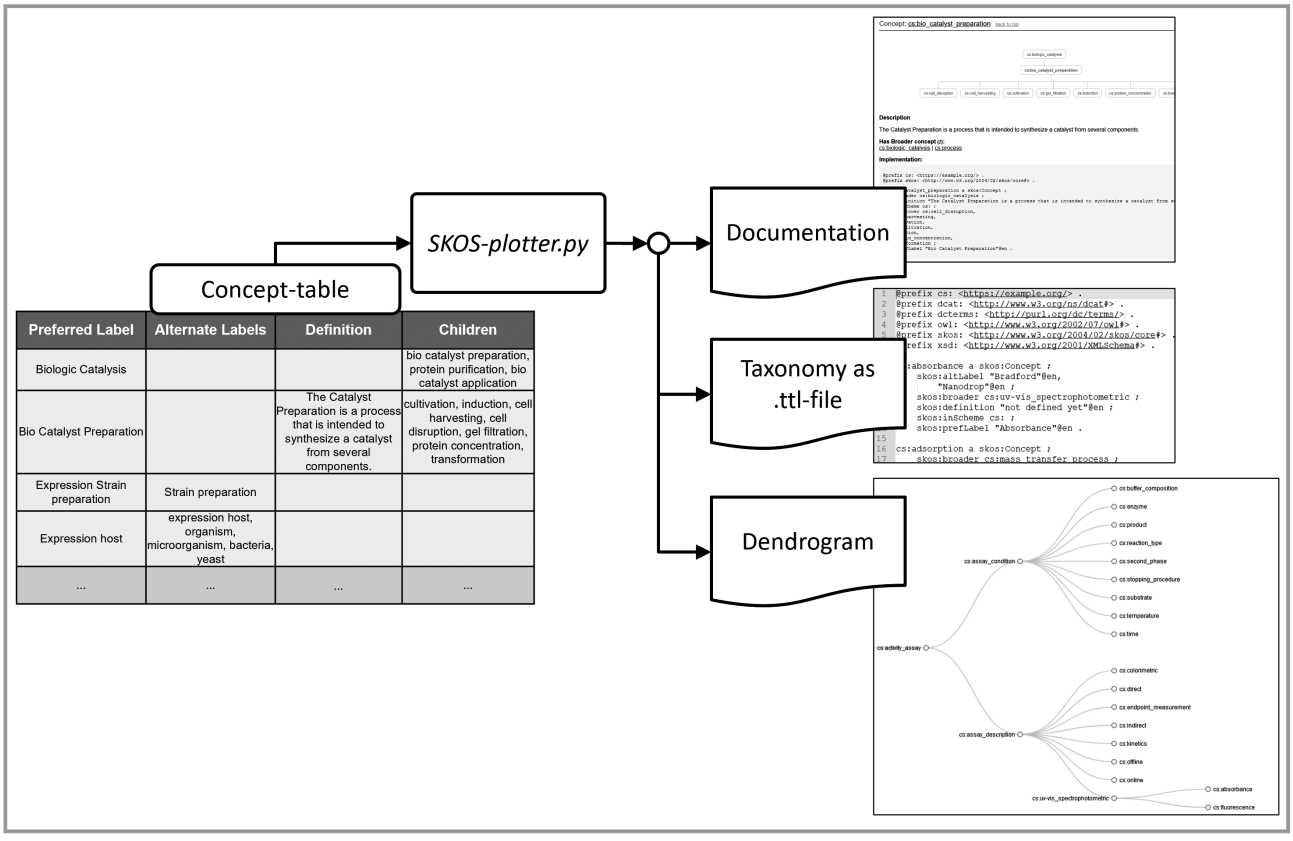

Figure 5. Input and output of the SKOS-plotter.py code, loading a concept-table provided by domain experts and yielding the taxonomy as turtle file, as well as dendrogram and documentation .html files.

 © 2022 The Authors. Chemie Ingenieur Technik published by Wiley-VCH GmbH

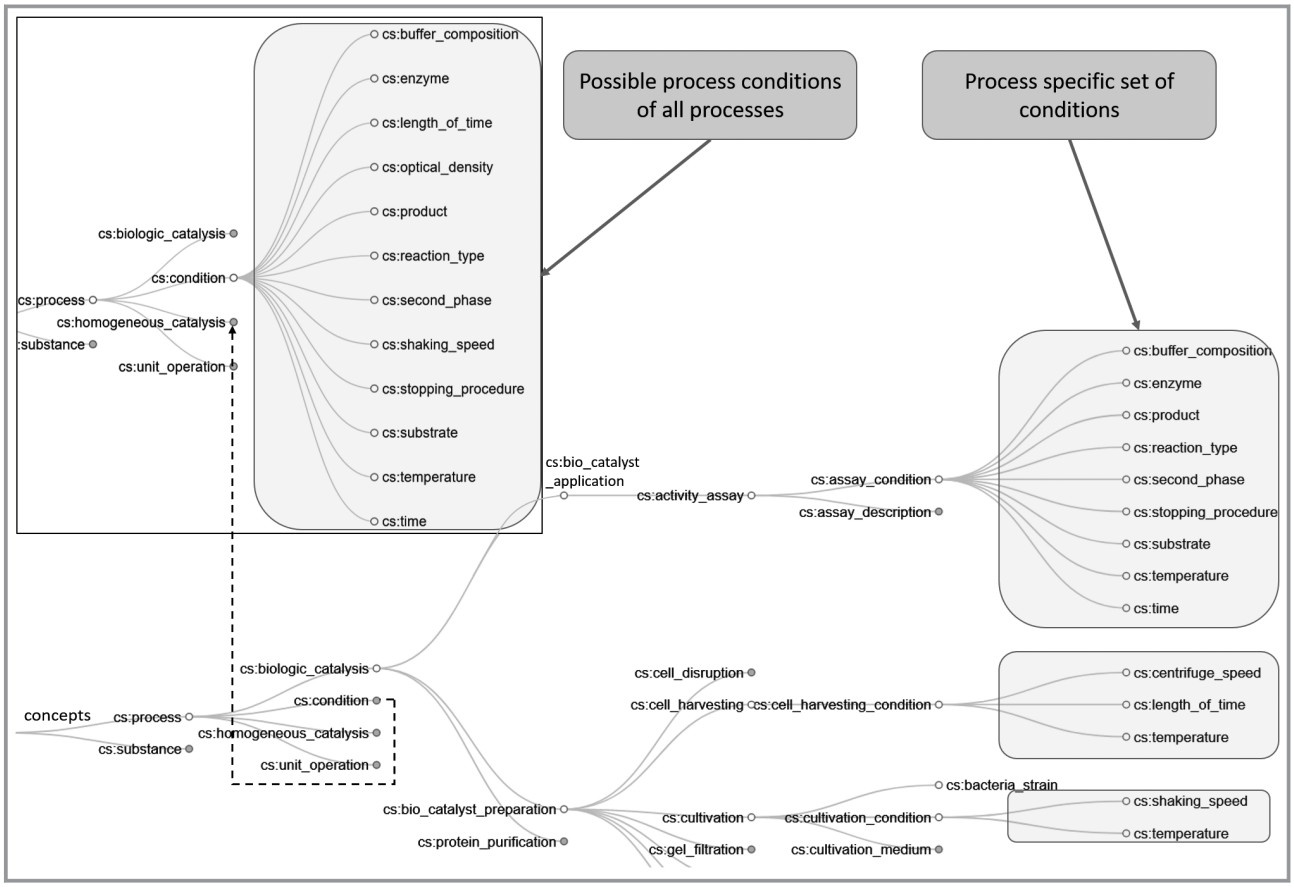

Figure 6. Part of overall class hierarchy depicted as dendrogram. The concept condition encompasses all possible process conditions of all processes (top left part of figure), while each process (e.g., activity assay, top right of figure) reuses the applicable concepts to a process-specific set of conditions.

3 Perspectives and Outlook

In the future, biocatalytic data can be associated with the semantic information of a working intermediate of the though developed biocatalysis ontology. This will enable much more elaborate queries on the data, like "give me the sequences of all transaminase activities with phenylethylamine as substrate" or "which substrates are transformed by pig liver esterase with a unit of $>1000\,\mathrm{min}^{-1}$ and mol?" Queries of this kind can be immediately done with, e.g., the SPARQL language which is supported by most ontology graph databases (triple or quad stores) since SPARQL query engines (SPARQL endpoints) are already included.

It will also enable machine learning algorithms to "understand" the underlying data and perform automated model building. For the researcher, data visualization will become much easier, since the visualization software can automatically choose the right columns in a data set to plot. For the enzymologist, data constancy checks can be performed, warning in case of data incompatibility or incomparable measurements.

These examples illustrate the powerful feature of semantic, ontology-based technology and many more applications will arise, once the technology is established.

4 Conclusion

In this article the first steps in direction of a standardized ontology for biocatalysis were described. We further developed open source, python-based tools to simplify this development process [48]. It is an iterative process with sub-optimal, but usable ontologies as intermediates. Currently a first draft of an initial collection of terms in form of a SKOS thesaurus is finalized, which will be published in the near future. The selection of the top-level ontology still needs to be decided – pragmatic considerations (BFO is currently used by more ontologies, but its root concepts originated from certain biologic experiments in mind, and this historic inheritance can still be seen) versus future prove clarity (EMMO is build on very well defined philosophical and physical terms and allows multi-perspectives, but is newer and therefore less frequently used). In further steps our SKOS thesaurus will be transformed into an ontology, using

 © 2022 The Authors. Chemie Ingenieur Technik published by Wiley-VCH GmbH

OWL and published in a git repository for versioning and ease of access.

There are still many steps to go and it will be required to include many international and interdisciplinary research groups working in the fields related to biocatalysis to form a widely accepted biocatalysis ontology.

We thank the Deutsche Forschungsgemeinschaft within the NFDI initiative (grant no: NFDI/2-1 - 2021) and the German Bundesministerium für Bildung und Forschung (BMBF) within the KIWI-biolab project (grant no: FKZ: 01DD20002C) for funding part of this work.
We further thank Peter Benner for carefully reading the manuscript. Open access funding enabled and organized by Projekt DEAL.

Abbreviations

FAIR	Findability, accessibility, interoperability, and reusability
HTML	Hypertext Markup Language
NFDI	Nationale Forschungsdateninfrastruktur (national research data infrastructure)
NFDI4CAT	Nationale Forschungsdateninfrastruktur for Catalysis-Related Sciences
OWL	Web Ontology Language
RDF	Resource Description Framework
SKOS	Simple Knowledge Organisation System
TA	task area
TTL	Terse RDF Triple Language
URI	Uniform Resource Identifier
W3C	World Wide Web Consortium
XML	Extensible Markup Language

References

[1] J. B. Pyser, S. Chakrabarty, E. O. Romero, A. R. H. Narayan, *ACS Cent. Sci.* **2021**, *7 (7)*, 1105–1116. DOI: https://doi.org/10.1021/acscentsci.1c00273
[2] S. Mazurenko, *ChemCatChem* **2020**, *12 (22)*, 5590–5598. DOI: https://doi.org/10.1002/cctc.202000933
[3] D. N. Woolfson, *J. Mol. Biol.* **2021**, *433 (20)*, 167160. DOI: https://doi.org/10.1016/j.jmb.2021.167160
[4] G. Linshiz, A. Goldberg, T. Konry, N. J. Hillson, *Perspect. Biol. Med.* **2012**, *55 (4)*, 503–520. DOI: https://doi.org/10.1353/pbm.2012.0044
[5] P. Halling, P. F. Fitzpatrick, F. M. Raushel, J. Rohwer, S. Schnell, U. Wittig, R. Wohlgemuth, C. Kettner, *Biophys. Chem.* **2018**, *242*, 22–27. DOI: https://doi.org/10.1016/j.bpc.2018.08.004
[6] M. Baker, *Nature* **2016**, *533*, 452–454. DOI: https://doi.org/10.1038/533452a
[7] A. G. McDonald, K. F. Tipton, *Molecules* **2022**, *27 (1)*, 263. DOI: https://doi.org/10.3390/molecules27010263
[8] K. F. Tipton, R. N. Armstrong, B. M. Bakker, A. Bairoch, A. Cornish-Bowden, P. J. Halling, J.-H. Hofmeyr, T. S. Leyh, C. Kettner, F. M. Raushel, J. Rohwer, D. Schomburg, C. Steinbeck, *Perspect. Sci.* **2014**, *1 (1–6)*, 131–137. DOI: https://doi.org/10.1016/j.pisc.2014.02.012
[9] S. Calvelage, M. Dörr, M. Höhne, U. T. Bornscheuer, *Adv. Synth. Catal.* **2017**, *359 (23)*, 4235–4243. DOI: https://doi.org/10.1002/adsc.201701079
[10] J. Pleiss, *ChemCatChem* **2021**, *13 (18)*, 3909–3913. DOI: https://doi.org/10.1002/cctc.202100822
[11] M. D. Wilkinson, M. Dumontier, Ij. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzales-Beltran, A. J. G. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. C't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, B. Mons, *Sci. Data.* **2016**, *3 (1)*, 160018. DOI: https://doi.org/10.1038/sdata.2016.18
[12] R. Zaru, J. Onwubiko, A. J. M. Ribeiro, K. Cochrane, J. D. Tyzack, V. Muthukrishnan, L. Pravda, J. M. Thornton, C. O'Donovan, S. Velanker, S. Orchard, A. Leach, M. J. Martin, *FEBS J.*, in press. DOI: https://doi.org/10.1111/febs.16168
[13] J. Range, C. Halupczok, J. Lohmann, N. Swainston, C. Kettner, F. T. Bergmann, A. Weidemann, U. Wittig, S. Schnell, J. Pleiss, *FEBS J.*, in press. DOI: https://doi.org/10.1111/febs.16318
[14] S. Malzacher, J. Range, C. Halupczok, J. Pleiss, D. Rother, *Chem. Ing. Tech.* **2020**, *92 (9)*, 1251–1251. DOI: https://doi.org/10.1002/cite.202055297
[15] V. Lapatas, M. Stefanidakis, R. C. Jimenez, A. Via, M. V. Schneider, *J. Biol. Res.-Thessalon.* **2015**, *22 (1)*, 9. DOI: https://doi.org/10.1186/s40709-015-0032-5
[16] D. Probst, M. Manica, Y. G. Nana Teukam, A. Castrogiovanni, F. Paratore, T. Laino, *Nat. Commun.* **2022**, *13 (1)*, 964. DOI: https://doi.org/10.1038/s41467-022-28536-w
[17] J. L. Faulon, M. du Lac, T. Duigou, P. Carbonell, J. Hérisson, Design Automation Workflows for Synthetic Biology and Metabolic Engineering: The Galaxy SynBioCAD portal, in *Proc. of the 12th International Workshop on Bio-Design Automation Worcester Polytechnic Institute*, Bio-Design Automation Consortium, Newton, MA **2020**, 25–26. www.iwbdaconf.org/2020/#proceedings
[18] P. S. F. Mendes, S. Siradze, L. Pirro, J. W. Thybaut, *ChemCatChem* **2021**, *13 (3)*, 836–850. DOI: https://doi.org/10.1002/cctc.202001132
[19] N. Hartl, E. Wössner, Y. Sure-Vetter, *Inform. Spektrum* **2021**, *44 (5)*, 370–373. DOI: https://doi.org/10.1007/s00287-021-01392-6
[20] C. Wulf, M. Beller, T. Boenisch, O. Deutschmann, S. Hanf, N. Kockmann, R. Kraehnert, M. Oezaslan, S. Palkovits, S. Schimmler, S. A. Schunk, K. Wagemann, D. Linke, *ChemCatChem* **2021**, *13 (14)*, 3223–3236. DOI: https://doi.org/10.1002/cctc.202001974
[21] www.w3.org/standards/semanticweb/data (Accessed on May 15, 2022)
[22] T. Decoene, B. De Paepe, J. Maertens, P. Coussement, G. Peters, S. L. De Maeseneire, M. De Mey, *Crit. Rev. Biotechnol.* **2018**, *38 (5)*, 647–656. DOI: https://doi.org/10.1080/07388551.2017.1380600
[23] *Handbook of Semantic Web Technologies* (Eds: J. Domingue, D. Fensel, J. A. Hendler), Springer, Berlin **2011**, 12.
[24] www.w3.org/standards/semanticweb/ (Accessed on May 15, 2022)

 © 2022 The Authors. Chemie Ingenieur Technik published by Wiley-VCH GmbH

[25] W. N. Borst, Construction of Engineering Ontologies for Knowledge Sharing and Reuse, Centre for Telematics and Information Technology (CTIT), Enschede **1997**.
[26] www.w3.org/OWL (Accessed on May 15, 2022)
[27] www.w3.org/TR/2012/REC-owl2-overview-20121211/ (Accessed on May 15, 2022)
[28] P. Strömert, J. Hunold, A. Castro, S. Neumann, O. Koepler, *Pure Appl. Chem.* **2022**. DOI: https://doi.org/10.1515/pac-2021-2007
[29] www.bioportal.bioontology.org/ontologies/ (Accessed on May 15, 2022)
[30] www.matportal.org/ontologies (Accessed on May 15, 2022)
[31] N. Guarino, *1st Int. Conf. on Formal Ontology in Information Systems, FOIS 1998*, Trento, Italy June **1998**.
[32] www.github.com/emmo-repo/ (Accessed on May 15, 2022)
[33] www.bioassayontology.org (Accessed on May 15, 2022)
[34] www.emmc.info/emmo-info/ (Accessed on May 15, 2022)
[35] www.github.com/emmo-repo/EMMO (Accessed on May 15, 2022)
[36] www.basic-formal-ontology.org/ (Accessed on May 15, 2022)
[37] S. Staab, R. Studer,~2nd ed., *Handbook on Ontologies*, Springer, Berlin 2009. DOI: https://doi.org/10.1007/978-3-540-92673-3
[38] www.w3.org/TR/skos-primer/ (Accessed on May 15, 2022)
[39] www.agrovoc.fao.org/browse/agrovoc/en/ (Accessed on May 15, 2022)
[40] www.w3.org/TR/skos-reference/ (Accessed on May 15, 2022)
[41] V. Haarslev, K. Hidde, R. Möller, M. Wessel, *Semantic Web* **2012**, *3 (3)*, 267–277. DOI: https://doi.org/10.3233/SW-2011-0032
[42] www.bitbucket.org/dtsarkov/factplusplus/src/master/ (Accessed on May 15, 2022)
[43] www.hermit-reasoner.com/ (Accessed on May 15, 2022)
[44] M.-O. Lenz, L. M. Ghiringhelli, C. Baldauf, M. Scheffler, Ontologies in Computational Materials Science, *Int. FAIR-DI Conf. on a FAIR Data Infrastructure for Materials Genomics*, June **2020**. https://th.fhi.mpg.de/meetings/fairdi2020/index.php?n=Meeting.PosterDetails&poster_id=18 (Accessed on May 15, 2022)
[45] https://github.com/surroundaustralia/VocExcel (Accessed on May 15, 2022)
[46] https://github.com/nfdi4cat/VocExcel (Accessed on May 15, 2022)
[47] https://github.com/lambdamusic/Ontospy (Accessed on May 15, 2022)
[48] https://github.com/TUDoAD/SKOS-plotter (Accessed on May 15, 2022)
[49] M. A. Musen, *AI Matters.* **2015**, *1 (4)*, 4–12. DOI: https://doi.org/10.1145/2757001.2757003

 © 2022 The Authors. Chemie Ingenieur Technik published by Wiley-VCH GmbH