1st Conference on Research Data Infrastructure

Harmonizing RDM

https://doi.org/10.52825/CoRDI.v1i.356

Published 07 Sept. 2023

Leveraging Terminology Services for FAIR Semantic Data Integration across NFDI Domains

How to Integrate Terminology Services Into Other Service Applications

Roman Baum[1][https://orcid.org/0000-0001-5246-9351] and Oliver Koepler[2][https://orcid.org/0000-0003-3385-4232]

[1] ZB MED - Information Centre for Life Sciences, Cologne, Germany

[2] TIB – Leibniz Information Centre for Science and Technology, Hannover, Germany

Abstract. The National Research Data Infrastructure (NFDI) strives to develop FAIR research data and data services for major scientific disciplines, using terminologies as a key factor for semantic annotations and semantic interoperability of data. Several NFDI consortia provide domain-specific terminologies through Terminology services or registries, offering access, search capabilities, visualization, and downloads. Prioritizing user-friendly access, terminology services seamlessly integrate semantic concepts into applications, often operating in the background to enable smooth semantic annotation and data interoperability. We present exemplary fields of application from selected disciplines and how terminology services support semantic search, user experience, annotation workflows, terminology curation and design.

Keywords: Terminology Services, FAIR metadata

1. Introduction

The National Research Data Infrastructure (NFDI) aims to create FAIR research data and data services for all major scientific disciplines. Terminologies are a critical success factor for describing research data. Terminologies are applied to define concepts in metadata schemas to enable consistent and structured descriptions of data and their relationships. Therefore, we are observing an increasing use of the application of terminologies across the NFDI consortia, i.e. their discussion and harmonization in the working groups of the NFDI section (Meta)data, Terminologies and Provenance. To support researchers, data stewards, ontology experts and other services, several NFDI consortia such as DataPLANT [1], NFDI4Biodiversity [2], NFDI4Chem [3], NFDI4Earth [4], NFDI4Health [5], NFDI4Ing [6], or NFDI4Objects [7] provide domain-specific terminologies in respective domain-specific terminology services or registries.

Use cases vary across the disciplines, but all of the services should be as user-friendly as possible when it comes to accessing and selecting concepts. In such a user-friendly approach, semantic concepts should be automatically integrated into an application, rather than requiring time-consuming manual input. In fact, many application areas are not so obvious. Terminology services are often used in the background, embedded in other services.

A terminology service could be the backbone of a research data infrastructure. There are several application areas, such as 1) semantic search, 2) user experience enhancement, or 3) annotation services, where terminology services can provide support.

2 Existing Terminology Services in the NFDI

Due to constraints of this format, we cannot fully describe all terminology services in all domains. Similar fields of application and approaches are valid for other services as well. We will focus on the two services SemLookP and the NFDI4Chem Terminology Service which are both based on the Ontology Lookup Service (OLS) [8].

2.1 SemLookP

SemLookP is a semantic lookup platform that provides terminologies from the fields of medicine, nutrition and life sciences. The access to these terminologies is realized via a graphical user interface (GUI) or via APIs. The GUI of SemLookP consists of several combined JavaScript based widgets. Such a widget uses the OLS API and combines the data received from the API with specific HTML components. These widgets can also be directly integrated into other applications.

2.2 NFDI4Chem Terminology Service

The NFDI4Chem Terminology Service provides access to a collection of ontologies relevant to the chemistry community. The collection was derived from an evaluation process [9] and is part of the Ontologies4Chem endeavor [10]. The Terminology Service provides faceted search inside ontologies, tree and lists views of concepts, properties and individuals and how to access them. The service not only provides an overview of ontologies in the domain, but also aims to support comparison and analysis across multiple ontologies for curation tasks. For this It offers an unified view of issues from the original ontology repositories within the Terminology GUI. The NFDI4Chem Terminology Service can also index and display SKOS vocabularies. The terminology service also provides a comprehensive API to retrieve all data and information and embed it into other services and applications.

3. Applications and Integration of Terminology Services

In the following we demonstrate some application and integration cases of terminology services in other applications mostly based on SemLookP and the NFDI4Chem Terminology Service. We will present 3 different topics where and how a terminology service could be easily integrated into an application.

3.1 Semantic Search

preVIEW is the first application we will focus on in this section. preVIEW is a user-friendly COVID19-related preprint viewer with advanced semantic search functionality. [11, 12] Since the terminologies from SemLookP are used for the annotation process, it is possible to use these concepts in the semantic search of preVIEW (see Figure 1).

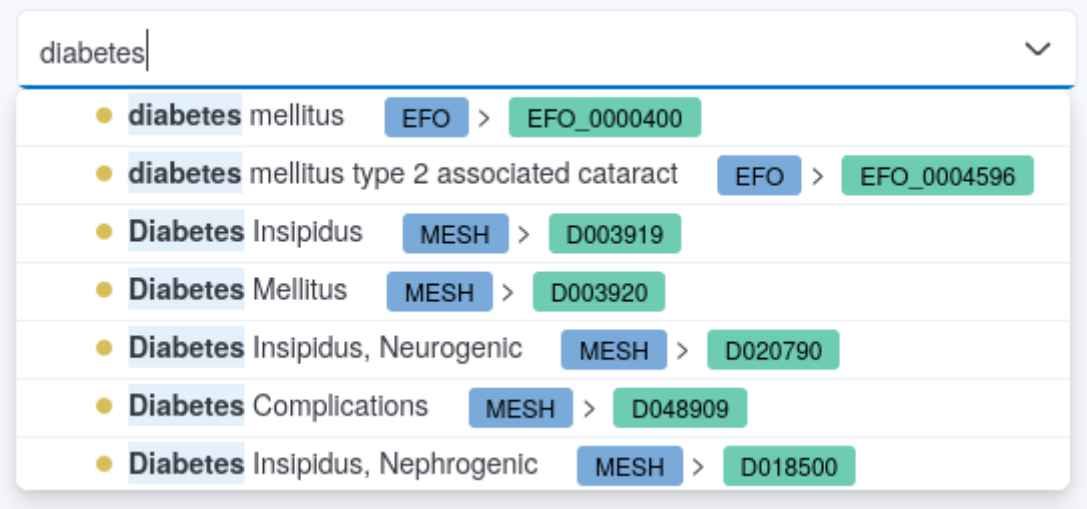

Figure 1. Autocomplete widget which is integrated into preVIEW.

There is also the GFBio Data Portal application. The GFBio Data Portal uses a RESTful API of the GFBio Terminology Service [13] for the annotation and the semantic search. Since the GFBio services are self-developed, it also proves that the integration of a terminology service into another application is possible with systems other than OLS.

3.2 User Experience Enhancement

To enhance the user experience, we will again focus on preVIEW. The annotations in preVIEW are highlighted with a background color. Clicking on an annotated concept opens a window. This window contains the metadata of the annotated concept (see Figure 2).

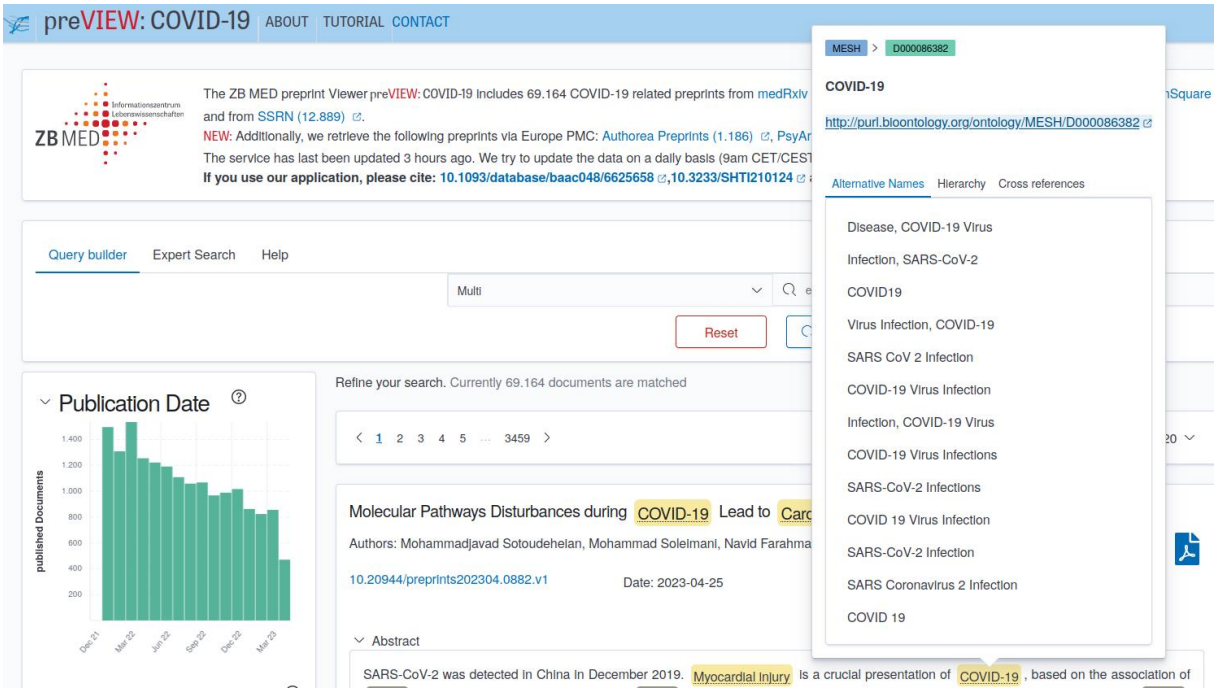

Figure 2. Metadata widget which is integrated in preVIEW.

Next to directly integrate widgets of a terminology service to another application it is also possible to receive the presented data via the API of a terminology service. This data could be used to develop custom HTML components. An example would be a semantic search filter. Imagine a user enters a specific term in a search bar. A hierarchical filter listing sub-

concepts could then appear next to the list of results. Such a sub-concept could act as a shortcut to further specify the search term.

3.3 Data Annotation Services

Many use cases of the NFDI4Chem Terminology Service can be described as service-to-service applications. The extensive API of the terminology service can be used by other services like the chemotion electronic lab notebook (ELN) or data repositories like RADAR4Chem, nmrXiv or others. The ELN reuses concepts from the named reaction ontology or chemical methods ontology to annotate reactions in experiments or analytical methods applied for the generation of data.

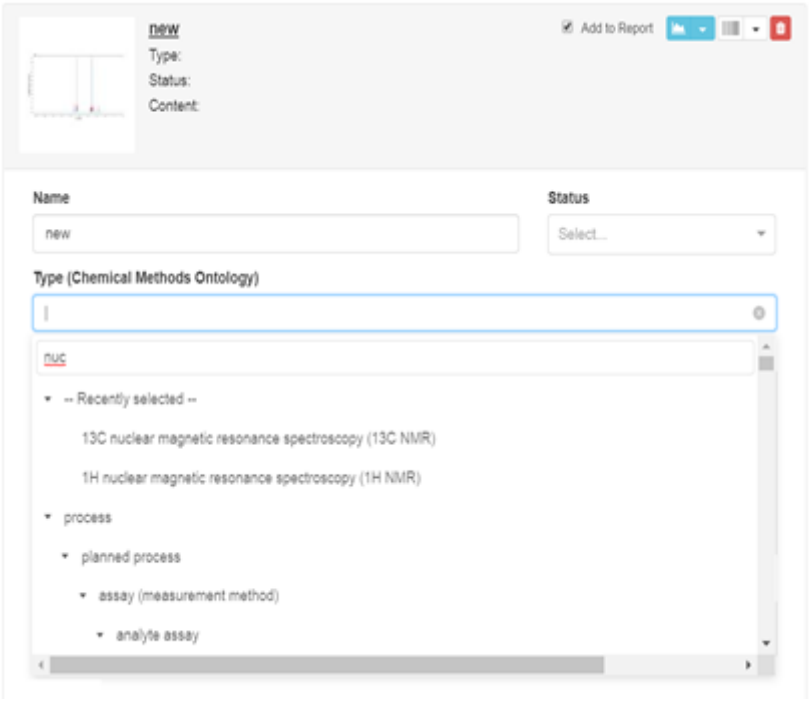

Figure 3. Reuse of CHMO terms in Chemotion ELN.

If not already provided by ELNs, repositories apply terms from ontologies for molecular entities, analytical methods, or processes to semantically annotate research data in their forms during the upload workflow.

This approach is freely transferable to other domains. For example, the NFDI4Health Metadata Annotation Workbench uses the API of SemLookP.

4. Conclusion and Outlook

In this work, we presented a very simple and smart solution for integrating terminology services into other applications by using the API or widgets.

However, domain-specific terminology services have limitations. These terminology services are isolated in their own domain. To solve this problem, terminology services could be connected and harmonized through an additional gateway in the future (see Figure 4). Such an approach would bridge the siloed solution and enable cross-domain fields of application.

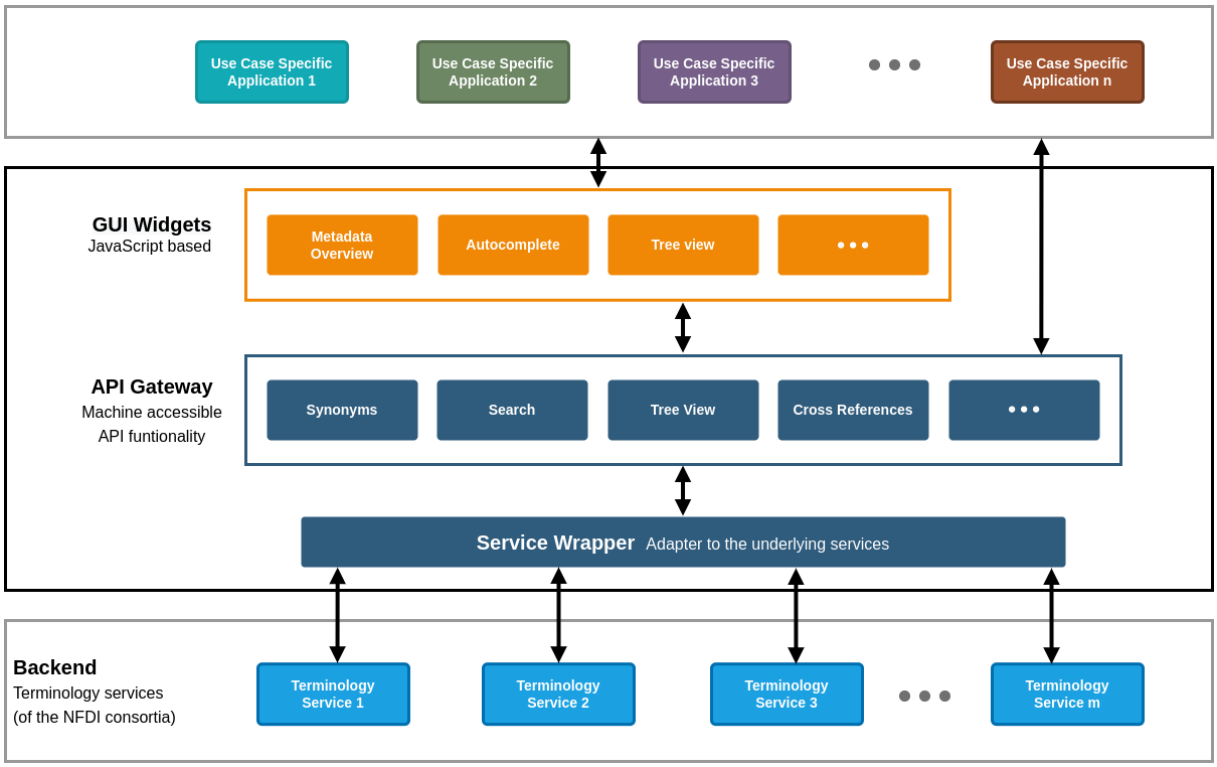

Figure 4. Architecture of a shared terminology API/widget tool set.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was done as part of the Consortia NFDI4Health and NFDI4Chem. We gratefully acknowledge the financial support of the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - grant numbers: 442326535 (NFDI4Health), 441958208 (NFDI4Chem).

References

1. SwateOntology DB / Swobup, https://nfdi4plants.org/content/service.html (last accessed 25 April 2023)
2. GFBio Terminology Service, https://terminologies.gfbio.org/ (last accessed 25 April 2023)
3. NFDI4Chem Terminology Service, https://terminology.nfdi4chem.de (last accessed 25 April 2023)
4. GEMET - GEneral Multilingual Environmental Thesaurus, https://www.eionet.europa.eu/gemet/ (last accessed 25 April 2023)
5. SemLookP, https://semanticlookup.zbmed.de/ (last accessed 25 April 2023)
6. NFDI4Ing Terminology Service, https://terminology.nfdi4ing.de/ts/ (last accessed 25 April 2023)
7. DANTE, https://api.dante.gbv.de/ (last accessed 25 April 2023)
8. S. Jupp, T. Burdett, C. Leroy, and H. Parkinson, "A new Ontology Lookup Service at EMBL-EBI", presented at the Workshop on Semantic Web Applications and Tools for Life Sciences, 2015. Accessed: Apr. 25, 2023. [Online]. Available:

https://www.semanticscholar.org/paper/A-new-Ontology-Lookup-Service-at-EMBL-EBI-Jupp-Burdett/b83bfbfc1f2f08e5b88af5ef65ef2a8687ac4112
9. P. Strömert, J. Hunold, A. Castro, S. Neumann, and O. Koepler, “Ontologies4Chem: the landscape of ontologies in chemistry,” Pure Appl. Chem., Mar. 2022, doi: https://doi.org/10.1515/pac-2021-2007.
10. P. Strömert, J. Hunold, and O. Koepler, “1st Ontologies4Chem Workshop – Ontologies for chemistry,” Sep. 07, 2022, doi: https://doi.org/10.25798/frnp-sn04.
11. L. Langnickel, R. Baum, J. Darms, S. Madan, and J. Fluck “COVID-19 preVIEW: Semantic Search to Explore COVID-19 Research Preprints,” Public Health and Informatics, vol.281, pp. 78-82, May, 2021, doi: https://doi.org/10.3233/SHTI210124
12. L. Langnickel, J. Darms, R. Baum, and J. Fluck “preVIEW: from a fast prototype towards a sustainable semantic search system for central access to COVID-19 preprints,” EAHIL, vol.17, no.3, pp. 8-14, Sep., 2021, doi: https://doi.org/10.32384/jeahil17484
13. N. Karam, C. Müller-Birn, M. Gleisberg et al., "A Terminology Service Supporting Semantic Annotation, Integration, Discovery and Analysis of Interdisciplinary Research Data," Datenbank Spektrum 16, pp 195–205, Oct. 05, 2016, doi: https://doi.org/10.1007/s13222-016-0231-8