Terminology service for research data management and knowledge discovery in low-temperature plasma physics

Authors:
Markus M. Becker[1]*, Ihda Chaerony Siffa[1], and Roman Baum[2]
*Lead presenter, markus.becker@inp-greifswald.de
[1]Leibniz Institute for Plasma Science and Technology (INP), Greifswald, Germany
[2]ZB MED - Information Centre for Life Sciences, Cologne, Germany

Abstract:
Terminology services (TS) [1,2] play a pivotal role in achieving structured metadata by providing controlled vocabularies and ontologies that standardize the description of data. This is a crucial aspect of research data management (RDM) in all scientific disciplines. In addition, TS facilitate the use of a common vocabulary within a scientific community also in a more general context, e.g. to annotate scientific papers, patents or other content for better discoverability, as envisaged by the Open Research Knowledge Graph (ORKG) [3] or the Patents4Science project [4].

To make use of these opportunities, terminologies, ontologies and knowledge graphs must be developed and made available as TS where they do not yet exist. This step is currently being taken by the research community in low-temperature plasma (LTP) physics. LTP physics explores partially ionized gases and its technological applications. This vibrant field offers innovative solutions for societal challenges, ranging from developing efficient lighting and solar cells to revolutionizing healthcare through plasma medicine. Various activities and projects have been started in the past years to support the RDM in LTP research and development and to facilitate the application of data-driven research methods. These activities are supported in parts by the NFDI4BIOIMAGE consortium, active work in the NFDI section “(Meta)data, Terminologies, Provenance”, and the basic service Terminology Services 4 NFDI (TS4NFDI) funded by Base4NFDI.

Recently, the ontology Plasma-O [5–7] for LTP physics has been developed at INP in collaboration with FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, providing a framework for structuring metadata and building a knowledge graph for scientific information within the field. The present contribution will show how a TS utilizing this resource can support different aspects of RDM and knowledge discovery using concrete examples. The application cases include (*i*) standardizing data annotation: By providing researchers with a controlled vocabulary of LTP-specific terms and their relationships, ensuring consistent and unambiguous data descriptions; (*ii*) enabling semantic search: Moving beyond keyword-based searches, TS allow for complex queries based on the relationships between concepts, significantly improving data discoverability; (*iii*) facilitating data integration: By mapping data from different sources to a common ontology, TS enable seamless integration and analysis of heterogeneous datasets, which is crucial for data-driven research and development. The TS Suite of TS4NFDI with the provided widgets [8] fits perfectly to the requirements of these three application cases and will support the harmonization of metadata in LTP physics. The implementation of a public TS is required to provide the domain-specific metadata in a standardized format and will be instrumental in unlocking the full potential of the TS widgets for RDM and knowledge discovery by LTP researchers.

Furthermore, the results can provide insights to other domains on how to apply TS to their specific needs.

The work was supported in parts by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the National Research Data Infrastructure – [NFDI46/1] – 501864659 and project number 496963457 as well as by the Federal Ministry of Education and Research (BMBF), project number 16KOA013A.

Keywords: terminology service, plasma physics, ontology, knowledge graph

References:

[1] S. Jupp, T. Burdett, C. Leroy, H. Parkinson, “A new Ontology Lookup Service at EMBL-EBI”, Workshop on Semantic Web Applications and Tools for Life Sciences (2015), https://ceur-ws.org/Vol-1546/paper_29.pdf (accessed: 2024-09-20).

[2] P. L. Whetzel, N. F. Noy, N. H. Shah, P. R. Alexander, C. Nyulas, T. Tudorache, M. A. Musen, “BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications”, *Nucleic Acids Res.* 39 (2011) W541–W545, https://doi.org/10.1093/nar/gkr469.

[3] Open Research Knowledge Graph, https://orkg.org/ (accessed: 2024-09-20).

[4] Patents4Science – Establishing an Information Infrastructure for the Use of Patent Knowledge in Science, https://www.patents4science.org/ (accessed: 2024-09-20).

[5] H. Sack, F. Hoppe, “Verbundprojekt: Qualitätssicherung und Vernetzung von Forschungsdaten in der Plasmatechnologie - QPTDat; Teilvorhaben: Wissensgraph und Ontologieentwicklung zur Vernetzung von Metadaten : Schlussbericht des Teilvorhabens”, 2023, https://doi.org/10.2314/KXP:1883436974.

[6] I. Chaerony Siffa, R. Wagner, L. Vilardell Scholten, M. M. Becker, “Semantic Information Management in Low-Temperature Plasma Science and Technology with VIVO”, 2024, preprint, https://doi.org/10.48550/arXiv.2409.11065.

[7] I. Chaerony Siffa, R. Wagner, L. Vilardell Scholten, M. M. Becker, “Plasma Ontology and Knowledge Graph Initial Release v0.5.0”, 2024, Zenodo, https://doi.org/10.5281/zenodo.13325226.

[8] J. Sasse, V. Kneip, R. Baum, P. Zimmermann, J. Darms, J. Schneider, V. Clemens, P. Oladazimi, L. Kühnel, “ts4nfdi/terminology-service-suite: v2.6.0”, 2024, Zenodo, https://doi.org/10.5281/zenodo.13692297.