2nd Conference on Research Data Infrastructure: RDM in Action

TIBconferencesession

eventual DOI

NFDI4DSO Version 2.0.0: A BFO Compliant Ontology for Data Science

Genet Asefa Gesese [1], Zongxiong Chen[2],
Heike Fliegl[1],
Sonja Schimmler[23], and
Harald Sack[14]

[1]FIZ Karlsruhe, Leibniz Institute for Information Infrastructure, Germany

[2]Fraunhofer FOKUS, Berlin, Germany

[3]Technical University of Berlin, Berlin, Germany

[4]Karlsruhe Institute of Technology, KIT, Germany
*Correspondence: Genet Asefa Gesese, genet-asefa.gesese@fiz-kalrsruhe.de

Abstract

Data Science (DS) is a multidisciplinary field that integrates mathematics, statistics, computer science and domain-specific knowledge to extract meaningful insights from diverse data sources, involving a variety of artifacts such as datasets, models, ontologies [1], code repositories, and execution platforms. The NFDI4DataScience (NFDI4DS) project aims to improve the FAIRness (Findable, Accessible, Interoperable, and Reusable) of research artifacts within the National Research Data Infrastructure (NFDI) framework. To achieve this, the initial NFDI4DS Ontology (NFDI4DSO Version 1.0.0) [2] was developed, based on the NFDICore Ontology Version 2.0 [3]. NFDI4DSO Version 1.0.0 primarily supports the Research Information Graph (RIG), which captures metadata about the resources, persons and organizations of the NFDI4DS consortium. In contrast, NFDI4DSO Version 2.0.0 significantly extends its focus beyond RIG by supporting the Research Data Graph (RDG), enabling the semantic representation and interlinking of diverse research data assets.

NFDI4DSO Version 2.0.0 is built upon NFDICore Version 3.0.0[1], which is mapped to the Basic Formal Ontology (BFO) [4] to enable broader interoperability. This enhanced mapping ensures seamless integration across different research domains. The NFDI4DSO Version 1.0.0 ontology has been successfully used to create the first instance of the NFDI4DS Knowledge Graph (NFDI4DS-KG), providing a structured and semantically rich representation of research information within the consortium. Furthermore, it served as the foundational schema for developing a named entity recognition dataset (NER)[2], to support downstream tasks such as information extraction and semantic annotation. Building on these applications, NFDI4DSO Version 2.0.0 is also planned to be utilized for similar purposes such as KG construction.

[1] https://ise-fizkarlsruhe.github.io/nfdicore/3.0.0/
[2] https://nfdi4ds.github.io/nslp2025/docs/readme2kg_shared_task.html

Resources

- NFDI4DSO Version 1.0.0 https.doi.org/ The initial version of the NFDI4DS ontology.
- NFDI4DS-KG: https://nfdi.fiz-karlsruhe.de/4ds/shmarql The NFDI4DS KG created based on the NFDI4DSO Version 1.0.0.

Author contributions

- **Conceptualization**: Genet Asefa Gesese
- **Software**: Genet Asefa Gesese, Zongxiong Chen
- **Supervision**: Sonja Schimmler, Harald Sack
- **Funding acquisition**: Sonja Schimmler
- **Methodology**: Genet Asefa Gesese, Zongxiong Chen
- **Writing – original draft**: Genet Asefa Gesese, Zongxiong Chen
- **Project administration**: Genet Asefa Gesese, Heike Fliegl
- **Writing – review & editing**: Sonja Schimmler, Harald Sack, Heike Fliegl

Competing interests

The authors declare that they have no competing interests.

Funding

This publication was written by the NFDI consortium NFDI4DataScience in the context of the work of the association German National Research Data Infrastructure (NFDI) e.V. NFDI is financed by the Federal Republic of Germany and the 16 federal states and funded by the Federal Ministry of Education and Research (BMBF) – funding code M532701 / the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - project number NFDI4DataScience (460234259).

References

[1] J. Chen, O. Mashkova, F. Zhapa-Camacho, R. Hoehndorf, Y. He, and I. Horrocks, “Ontology embedding: A survey of methods, applications and resources,” *IEEE Transactions on Knowledge and Data Engineering*, 2025.

[2] G. A. Gesese, J. Waitelonis, Z. Chen, S. Schimmler, and H. Sack, “Nfdi4dso: Towards a bfo compliant ontology for data science,” *arXiv preprint arXiv:2408.08698*, 2024.

[3] O. Bruns, T. Tietz, J. Waitelonis, E. Posthumus, and H. Sack, “Nfdicore 2.0: A bfo-compliant ontology for multi-domain research infrastructures,” *arXiv preprint arXiv:2410.01821*, 2024.

[4] J. N. Otte, J. Beverley, and A. Ruttenberg, “Bfo: Basic formal ontology,” *Applied ontology*, vol. 17, no. 1, pp. 17–43, 2022.