How are NFDI consortia using Knowledge Graphs?

An overview of common functions and challenges by the Working Group "Knowledge Graphs"

Lozana Rossenova[1] https://orcid.org/0000-0002-5190-1867, Fidan Limani[2] https://orcid.org/0000-0002-5835-2784,
Carsten Fortmann-Grote[3] https://orcid.org/0000-0002-2579-5546,
Angelika Kaplan[4] https://orcid.org/0009-0009-9101-5833, Florian Thiery[5] https://orcid.org/0000-0002-3246-3531,
Atif Latif[2] https://orcid.org/0000-0003-3085-3031, Renat Shigapov[6] https://orcid.org/0000-0002-0331-2558,
Benjamin Zapilko[7] https://orcid.org/0000-0001-9495-040X, Moritz Schubotz[8] https://orcid.org/0000-0001-7141-4997,
Heike Fliegl[8] https://orcid.org/0000-0002-7541-115X, Tabea Tietz[8] https://orcid.org/0000-0002-1648-1684

[1] TIB – Leibniz Information Centre for Science and Technology

[2] ZBW – Leibniz Informationszentrum Wirtschaft

[3] Max Planck Institute for Evolutionary Biology

[4] Karlsruhe Institute of Technology (KIT)

[5] Leibniz-Zentrum für Archäologie (LEIZA), Research Squirrel Engineers Network

[6] University Library, University of Mannheim

[7] GESIS – Leibniz Institute for the Social Sciences

[8] FIZ Karlsruhe – Leibniz Institut für Informationsinfrastruktur GmbH

Abstract

The NFDI Working Group "Knowledge Graphs" (WG KGs) operates within the cross-cutting Section Metadata to coordinate and facilitate exchange on best practices, use cases, and challenges in the creation and use of knowledge graphs (KGs) across different disciplines. In this poster, we follow up on previously published findings, such as the dataset of KGs and related metadata developed and used by different consortia [1]. Besides the successful proposal and initiation of a basic service for a KG infrastructure (KGI4NFDI) [2], members of the WG have continued to meet regularly to discuss the ways the basic service can best meet the requirements of individual disciplines, and the NFDI as a whole. Furthermore, the WG discusses concerns beyond the scope of the basic service, such as efforts within individual consortia focused on disciplinary tasks, or ongoing developments within related basic services, such as TS4NFDI [3] or PID4NFDI [4]. Based on agile requirements gathering, feedback, and practical exchange, the WG has coordinated the collection of *functional use cases* and ongoing *challenges* different consortia face, with the aim to identify commonalities and/or distinctions that can inform the development of current and future basic services.

The core premise of KGs is to represent (meta)data and their relations in a machine-understandable format, which improves semantic interoperability and data integration by adopting standard ontologies and Linked Data principles (e.g., RDF, OWL). This drives the functional use cases observed in NFDI consortia. Following a literature review, the following core KG functionalities were categorised by members of the WG: a) Data integration; b) Data FAIRification; c) Metadata cataloguing; d) Data verification,

enrichment and extension; e) Machine actionability and AI-readiness [5–9]. These core categories were specified further through an open call among the consortia represented in the WG KG. The call collected an extended dataset of use case examples from their KG projects, grouped according to functionality categories, and linked to the previously published KG overview dataset [10]. Examples include KGs established by individual consortia functioning as data and metadata catalogues for discipline-specific repositories and research data portals, while using dedicated semantic ontologies as in the case of NFDI4Culture, NFDI-MatWerk, NFDI4DataScience, MaRDI, and NFDI4Objects, among others. A more specific example of data integration is the capacity of the Semantic Kompakkt KG (NFDI4Culture) and the NFDI4Bioimage KG to semantically express the annotation of 3D models and Bioimaging file formats, respectively.

Alongside the use cases that illustrate why different consortia adopt specific KG functionalities, common challenges were also identified and documented. These challenges include harmonizing diverse ontologies within or across consortia, ensuring data verification and quality, optimizing AI-ready tools to support KG (and data) (re)usage, as well as socio-technical challenges, such as data privacy, or managing access restrictions. By collecting concrete use case examples and identifying specific challenges, the WG KGs supports its mission to encourage broader adoption of KGs across the NFDI. This effort also contributes to building a "one-stop-shop" catalog of KG example implementations that the NFDI community can turn to when starting KG-driven projects.

Keywords: Knowledge Graphs, FAIR data, NFDI, semantic interoperability, data integration

Resources

- https://doi.org/10.5281/zenodo.15224731: The first draft of the KG dataset containing use cases and challenges collected from the NFDI community.

- https://kgi.services.base4nfdi.de/kg_registry/: KGI4NFDI registry containing metadata on the existing NFDI KGs.

Author contributions

All authors contributed to conceptualization and writing of this manuscript.

Competing interests

The authors declare that they have no competing interests.

Funding

Acknowledgement

References

[1] F. Limani, et al., "Who is using Knowledge Graphs in NFDI? An overview by the Working Group "Knowledge Graphs",' Zenodo, Sept. 10, 2023, doi: 10.5281/zenodo.8332776.

[2] L. Rossenova, R. Shigapov, M. Schubotz, F. Limani, B. Zapilko, and K. U. Förstner, "KGI4NFDI - Knowledge Graph Infrastructure for the German National Research Data Infrastructure," Zenodo, Jul. 28, 2024, doi: 10.5281/zenodo.13118749.

[3] O. Koepler, J. Sasse, N. Karamand, and R. Baum, "Terminology Services 4 NFDI (TS4NFDI) - Initialisation Phase Proposal," Zenodo, Jun. 20, 2024, doi: 10.5281/zenodo.12188787.

[4] R. Bertelmann, "PID4NFDI – Persistent Identifier Services for the German National Research Data Infrastructure: Proposal for the Initialisation Phase of Base4NFDI," Zenodo, Dec. 18, 2024, doi: 10.5281/zenodo.14281250.

[5] G. Sanou, et al., "IMGT-KG: A Knowledge Graph for Immunogenetics," in SEMANTIC WEB - ISWC 2022 (eds. Sattler, U. et al.), vol. 13489, pp.628–642, ChinaScope; Elsevier; IBM Res; Metaphacts; MEMECT; Oracle; Springer, 2022.

[6] J. Frey, D. Streitmatter, N. Arndt, and S. Hellmann, 'Reproducibility Crisis in the LOD Cloud? Studying the Impact of Ontology Accessibility and Archiving as a Counter Measure', in SEMANTIC WEB - ISWC 2022 (eds. Sattler, U. et al.), vol. 13489, pp.91–107, ChinaScope; Elsevier; IBM Res; Metaphacts; MEMECT; Oracle; Springer, 2022.

[7] D. Dessí, F. Osborne, D. Reforgiato Recupero, D. Buscaldi, and E. Motta, "CS-KG: A Large-Scale Knowledge Graph of Research Entities and Claims in Computer Science," in SEMANTIC WEB - ISWC 2022 (eds. Sattler, U. et al.), vol 13489, Springer, Cham, doi: 10.1007/978-3-031-19433-7_39.

[8] T. Tietz, O. Bruns, H. Fliegl, E. Posthumus, T. Schrade, and H. Sack, "Knowledge Graph-basierte Forschungsdatenintegration in NFDI4Culture," in Proc. of DHd 2023, Open Humanities Open Culture, 9. Tagung des Verbands Digital Humanities im deutschsprachigen Raum, 2023, doi: 10.5281/zenodo.7748740.

[9] M. Hofer, D. Obraczka, A. Saeedi, H. Köpcke, and E. Rahm, "Construction of Knowledge Graphs: Current State and Challenges," Information, 15(8), 509, 2024, doi: 10.3390/info15080509.

[10] R. Shigapov et al., "Dataset: an overview of knowledge graphs in NFDI," Zenodo, Jul. 07, 2023, doi: 10.5281/zenodo.8124286.