Knowledge Graph based RDM Solutions
NFDI4Culture - NFDI-MatWerk - NFDI4DataScience

Harald Sack[1][2][https://orcid.org/0000-0001-7069-9804], Torsten Schrade[3][https://orcid.org/0000-0002-0953-2818], Oleksandra Bruns[1][2][https://orcid.org/0000-0002-8501-6700], Etienne Posthumus[1][https://orcid.org/0000-0002-0006-7542], Tabea Tietz[1][2][https://orcid.org/0000-0002-1648-1684], Ebrahim Norouzi[1][2][https://orcid.org/xxx], Jörg Waitelonis[1][https://orcid.org/0000-0001-7192-7143], Heike Fliegl[1][https://orcid.org/0000-0002-7541-115X], Linnaea Söhn[3][https://orcid.org/0000-0001-8341-1187], Julia Tolksdorf[3][https://orcid.org/0000-0002-0495-5897], Jonatan Jalle Steller[3][https://orcid.org/0000-0002-5101-5275], Abril Azócar Guzmán[4][https://orcid.org/0000-0001-7564-7990], Said Fathalla[4][https://orcid.org/0000-0002-2818-5890], Ahmad Zainul Ihsan[4][https://orcid.org/0000-0002-1008-4530], Volker Hofmann[4][https://orcid.org/0000-0002-5149-603X], Stefan Sandfeld[4][https://orcid.org/0000-0001-9560-4728], Felix Fritzen[5][https://orcid.org/0000-0003-4926-0068], Amir Laadhar[5][https://orcid.org/0000-0001-9106-8825], Sonja Schimmler[6][https://orcid.org/0000-0002-8786-7250], and Peter Mutschke[7][https://orcid.org/0000-0003-3517-8071]

[1]FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany

[2]Karlsruhe Institute of Technology (AIFB), Kaiserstr. 89, 76133 Karlsruhe, Germany

[3]Academy of Sciences and Literature Mainz, Geschwister-Scholl-Straße 2, 55131 Mainz, Germany

[4]Institute for Advanced Simulations – Materials Data Science and Informatics (IAS-9), Forschungszentrum Jülich GmbH, Germany

[5]University of Stuttgart, Cluster of Excellence SimTec, Stuttgart Center for Simulation Science, Universitätsstraße 32 70569 Stuttgart, Germany

[6]Fraunhofer Institute for Open Communication Systems (FOKUS), Kaiserin-Augusta-Allee 31, 10589 Berlin, Germany

[7]GESIS - Leibniz Institute for the Social Sciences, Unter Sachsenhausen 6-8, 50667 Köln, Germany

Abstract: Based on our experience within the NFDI4Culture and NFDI-MatWerk projects we propose generalized knowledge graph based research data management solutions, which are applicable to other consortia. Our solution covers the construction of a common NFDI core ontology adapted to specific domains via domain extensions as a basis for a knowledge graph (KG) providing information about a consortium and its related research data and software resources. This KG serves as a backend for the web portal that enables interactive access and management of this data. Already implemented for NFDI4Culture and to be adapted by NFDI-MatWerk, this solution might serve as an example solution also for other consortia. We are synchronizing our efforts with ongoing work to implement knowledge graph based research data management in NFDI4DataScience.

Keywords: Semantics, Knowledge Graphs, Ontology
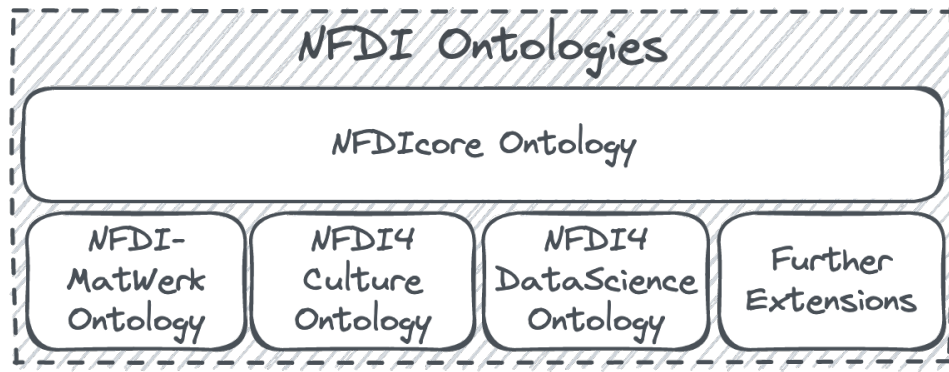
Figure 1. Schematic overview of the anticipated modular approach covering the NFDI core ontology and domain specific modular extensions.

1 NFDI Ontologies

Despite covering different scientific domains all NFDI consortia share similar concepts such as structure, organization, people, institutions, areas of expertise, data repositories, devices, infrastructure and much more that have to be represented when aiming at a semantic description [1]–[3]. This suggested the development of an NFDI core ontology to increase the interoperability of concepts across multiple NFDI consortia in the best sense of the FAIR Data Principles. A starting point was the NFDI4Culture ontology. However, the adoption by NFDI-MatWerk as a second consortium required several specific extensions, which made us realize that a modular approach works best for taking domain specific requirements into account. Figure 1 gives a schematic overview of the NFDIcore ontology, covering a consortium wide shared structure and domain specific extensions (modules) like NFDI4Culture (cto) [1], NFDI-MatWerk (mwo)[2] that are only relevant for a specific domain, and the NFDI4DataScience ontology that has a more interdisciplinary focus.

NFDIcore version 1.1 consists of 36 classes and 60 object attributes[3]. The NFDIcore classes have been linked to 24 existing ontologies, including frapo[4], fabio[5], void[6] and schema[7], following best practices in ontology development to ensure high semantic expressivity and interoperability. The ontology has been publicly available since June 2022, fully documented and integrated into the NFDI4Culture Information Portal[8].

2 The NFDI4Culture and NFDI-MatWerk Knowledge Graphs

We distinguish between two types of NFDI Knowledge Graphs (KGs): the Research Information Graph (RIG), covering metadata about the consortium's resources, persons, and organisations (aligned to the Common European Research Information Standard, CERIF[9]), and the Research Data Graph (RDG), covering content related index data from the consortium's heterogeneous data resources. The goal of the RIG is to en-

---

[1] https://github.com/ISE-FIZKarlsruhe/nfdi4culture-ontology

[2] https://nfdi-matwerk.pages.rwth-aachen.de/ta-oms/mwo/doc/index.html

[3] NFDIcore interactive view: https://service.tib.eu/webvowl/#iri=https://nfdi4culture.de/ontology.ttl

[4] https://sparontologies.github.io/frapo/current/frapo.html

[5] https://sparontologies.github.io/fabio/current/fabio.html

[6] http://vocab.deri.ie/

[7] https://schema.org/

[8] https://nfdi4culture.de/ontology.html
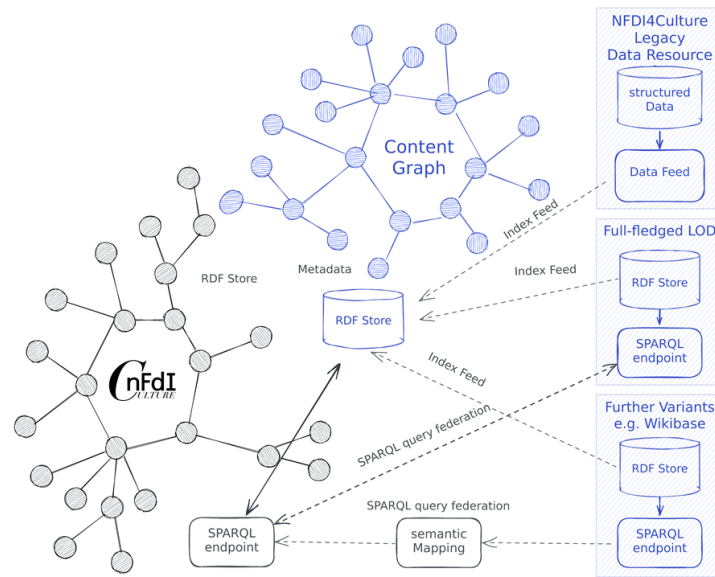
[9] https://eurocris.org/services/main-features-cerif

Figure 2. Overview of different data integrations. CC-BY 4.0, Authors: O. Bruns, E. Posthumus, T. Tietz, H. Sack

able exploration and retrieval of index and metadata pertaining to data resources, data services, persons, and institutions involved in NFDI consortia, while the RDG's primary objective is to facilitate access and exploration of content within the NFDI resources, as well as interconnection of the content from different resources within and across particular domains following the FAIR[4] principles. For fully fledged Linked Open Data (LOD) resources, SPARQL query federation enables access and cross connections between distributed resources. However, for reasons of efficiency and to enable across connections already locally in the RDG, the establishment of an enriched index containing metadata about entities was decided. To enable the inclusion of heterogeneous legacy data resources, including structured data and Wikibase-based resources, a metadata index harvesting mechanism has been designed based on a lightweight interchange format, the Graph Interchange Format (GIF), which so far has been implemented in a first use case as standalone protocol for NFDI4Culture as CGIF (Culture Graph Interchanged Format)[10]. (C)GIF can be embedded directly into any webpage to be extracted as RDF via a URL. Alternatively, data contributors can submit a (C)GIF resource in any RDF-compatible format. (C)GIF is designed to enable domain experts as contributors to easily contribute their data to the RDG without the need to implement complex APIs.

Figure 2 illustrates the concept of RIG and RDG, forming in sum the current NFDI4Culture KG[11] and the relevant data integration variants that are currently implemented. Further information is given in [5][12] and [6]. The NFDI-MatWerk KG[13] is set up in a similar manner and already allows simple retrieval via a SPARQL endpoint[14]. However, the (C)GIF will need to be adapted to the specific requirements of the Material Science Engineering domain.
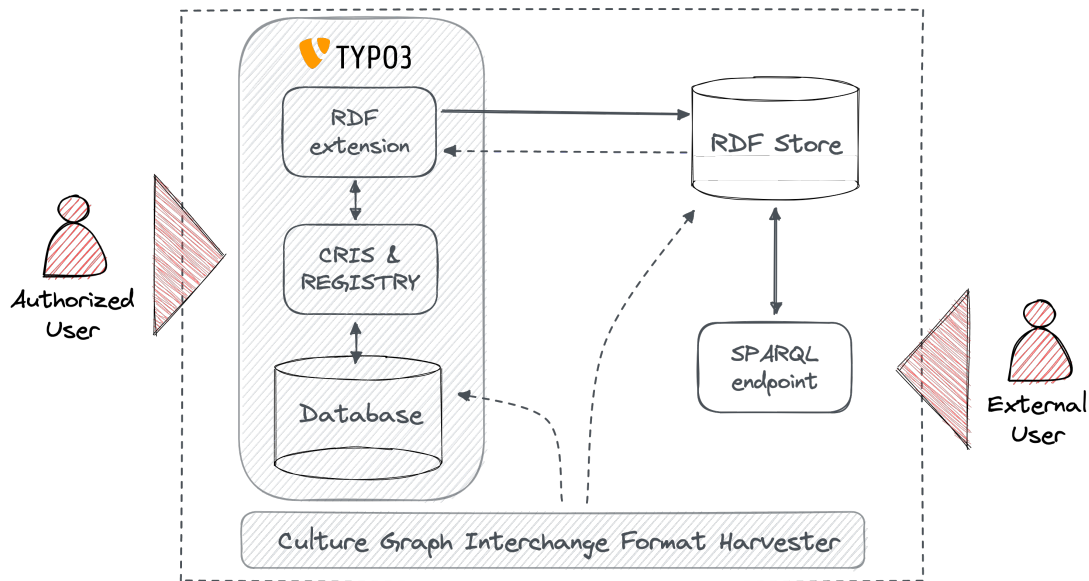
---

[10] https://docs.nfdi4culture.de/ta5-cgif-specification

[11] https://nfdi4culture.de/resources/knowledge-graph.html

[12] https://doi.org/10.5281/zenodo.7748740

[13] https://demo.fiz-karlsruhe.de/matwerk/

[14] https://demo.fiz-karlsruhe.de/sparql/

Figure 3. Overview of the Culture Information Portal Architecture. CC-BY 4.0, Author: T. Schrade

3 The NFDI4Culture Portal

NFDI4Culture is a pioneer in creating an information portal, providing a centralized access point to decentralized research data. The Culture Information Portal[15] serves as a user-friendly web-based Current Research Information System (CRIS). It enables a unified access point to the research data, services, etc. of the NFDI4Culture community. The portal adheres to international standards for CRIS systems and is implemented using the TYPO3 extensions "Academy Current Research Information System"[16] and "Linked Data for TYPO3"[17]. It allows non-expert users to contribute resources and metadata. The CRIS data is made available as Linked Open Data, with various RDF serializations and a standardized LOD API[18] based on the Hydra vocabulary[19]. Data is curated and continuously expanded in a decentralized manner for the Culture KG using ingest routines and the Oxigraph[20] native RDF store. The TYPO3 implementation ensures long-term sustainability through TYPO3 LTS releases, the freedom to integrate external ontologies, and use dedicated NFDI ontologies. The establishment of the Culture Information Portal not only highlights the consortium's innovative approach but also serves as a valuable model for other consortia seeking to create similar portals and improve access to their research data.

4 Future Work

The proposed plans for future work can be divided into three main areas: Firstly, there is a need for the continued development and widespread adoption of the NFDIcore ontology and KGs for both NFDI4Culture and NFDI-MatWerk, as well as NFDI4DataScience, NFDI4Memory, and other interested consortia. Secondly, it is essential to adapt and integrate ontologies and KGs into the existing infrastructure of NFDI4DataScience.

---

[15] https://nfdi4culture.de/
[16] https://github.com/digicademy/academy
[17] https://github.com/digicademy/lod
[18] https://nfdi4culture.de/resource/about.html
[19] https://www.hydra-cg.com/spec/latest/core/
[20] https://github.com/oxigraph/oxigraph

Lastly, a federated approach to accessing research data across multiple consortia is also being considered.

As the development of the proposed KG-based RDM infrastructure is still in its early phase, the current focus is on further developing the KG-based infrastructure and (C)GIF exchange format in NFDI4Culture and NFDI-MatWerk. Additionally, broad adoption of (C)GIF by NFDI4Culture participants, including further data repository variants such as infrastructures based on WissKI[21], will be fostered through hands-on tutorials within the consortium, workshops for the entire domain-specific community, and the publication of guidelines and best practices. The next step is to adopt the proposed ontology and KG architecture in NFDI4DataScience, NFDI4Memory, and beyond.

Our long term goal is to enable and explore inter domain connections of KGs across multiple consortia to enable FAIR access to research data over multiple scientific domains as one of the original visions of NFDI. The proposed NFDI core ontology is seen as a key to bring this vision a step forward.

Competing interests

The authors declare that they have no competing interests.

Funding

References

[1] I. Anders, T. Arera-Rütenik, S. Arndt, *et al.*, *Ontology Harmonization and Mapping - Working Group Charter (NFDI section-metadata)*, version 1.0, Jun. 2022. DOI: 10.5281/zenodo.6726519. [Online]. Available: https://doi.org/10.5281/zenodo.6726519.

[2] I. Anders, K. Bailly, R. Baum, *et al.*, *Terminology Services - Working Group Charter (NFDI section-metadata)*, version 1.0, Jun. 2022. DOI: 10.5281/zenodo.6759325. [Online]. Available: https://doi.org/10.5281/zenodo.6759325.

[3] M. Stocker, L. Rossenova, R. Shigapov, *et al.*, *Knowledge Graphs - Working Group Charter (NFDI section-metadata)*, version 1.2, Apr. 2023. DOI: 10.5281/zenodo.7802304. [Online]. Available: https://doi.org/10.5281/zenodo.7802304.

[4] E. Schultes and P. Wittenburg, "FAIR Principles and Digital Objects: Accelerating convergence on a data infrastructure," in *Data Analytics and Management in Data Intensive Domains: 20th Int. Conference, DAMDID/RCDL 2018*, Springer, 2019, pp. 3–16.

[5] T. Tietz, O. Bruns, H. Fliegl, E. Posthumus, T. Schrade, and H. Sack, "Knowledge Graph-basierte Forschungsdatenintegration in NFDI4Culture.," in *DHd2023: Open Humanities, Open Culture*, A. Busch and P. Trilcke, Eds., 2023.

[6] T. Tietz, O. Bruns, L. Söhn, *et al.*, "From Floppy Disks to 5-Star LOD: FAIR Research Infrastructure for NFDI4Culture," in *3rd Workshop on Metadata and Research (objects) Management for Linked Open Science (DaMaLOS), co-located with ESWC 2023*, L. J. Castro, D. Dessi, J. Dierkes, D. Rebholz-Schuhmann, and S. Schimmler, Eds., Publisso, 2023. DOI: 10.4126/FRL01-006444986. [Online]. Available: https://frl.publisso.de/resource/frl:6444986.

---

[21] https://wiss-ki.eu/