nfdi • Nationale Forschungsdaten Infrastruktur

Section
(Meta)data, Terminology, Provenance

Working Group Charter
Knowledge Graphs

Name of the working group

Knowledge Graphs

Acronym

section-metadata-wg-kg

Contact persons

Renat Shigapov (Universitätsbibliothek Mannheim, BERD@NFDI),
renat.shigapov@uni-mannheim.de

Lozana Rossenova (TIB – Leibniz-Informationszentrum Technik und
Naturwissenschaften und Universitätsbibliothek, NFDI4Culture),
lozana.rossenova@tib.eu

Moritz Schubotz (FIZ Karlsruhe - Leibniz Institute for Information Infrastructure,
MaRDI), moritz.schubotz@fiz-karlsruhe.de

Authors

Markus Stocker (0000-0001-5492-3212), Lozana Rossenova
(0000-0002-5190-1867), Renat Shigapov (0000-0002-0331-2558), Noemi Betancort
(0000-0002-0156-3556), Stefan Dietze, Bridget Murphy (0000-0002-1354-2381),
Christian Bölling (0000-0002-6544-1363), Moritz Schubotz (0000-0001-7141-4997),
Oliver Koepler (0000-0003-3385-4232)

Abstract

Knowledge Graphs are a key technology for implementing the FAIR principles in data infrastructures by ensuring interoperability for both humans and machines. The Working Group "Knowledge Graphs" in Section "(Meta)data, Terminologies, Provenance" of the German National Research Data Infrastructure (Nationale Forschungsdateninfrastruktur (NFDI) e.V.) aims to promote the use of knowledge graphs in all NFDI consortia, to facilitate cross-domain data interlinking and federation following the FAIR principles, and to contribute to the joint development of tools and technologies that enable transformation of structured and unstructured data into semantically reusable knowledge across different domains.

1. Motivation

Knowledge Graphs (KG) are a widely used database technology for the flexible description of entities and their linking enabling advanced information modeling, exploration, and use. Since the establishment of the Google Knowledge Graph[1], these technologies have gained momentum in industry and in research, where KGs are used to organize and publish information about both physical entities and information artifacts, like articles, datasets, software, samples, instruments, contributors, organizations and their interrelations in a machine-interpretable way. OpenAIRE Research Graph[2], GESIS Knowledge Graph Infrastructure[3], TIB Open Research Knowledge Graph[4], Springer Nature SciGraph[5], Research Graph Foundation[6], the GND.network[7], and PID Graph[8] are established infrastructures which connect research data and research-related entities and make it accessible. Besides this, the Wikidata knowledge graph[9] contains a large number of research-related entities[10] and several domain specific Wikibase[11] instances are used by various research communities[12].

Individual NFDI consortia develop services for the discovery of their assets, e.g., datasets, services, organizations, etc. For the research community served by a particular NFDI consortium, these services act as the primary entry point. However, given the increasingly

---

[1] https://blog.google/products/search/introducing-knowledge-graph-things-not
[2] https://graph.openaire.eu
[3] https://www.gesis.org/forschung/angewandte-informatik/knowledge-graph-infrastruktur
[4] https://orkg.org
[5] https://www.springernature.com/gp/researchers/scigraph
[6] https://researchgraph.org
[7] https://gnd.network
[8] https://commons.datacite.org
[9] https://www.wikidata.org
[10] https://scholia.toolforge.org
[11] https://wikiba.se
[12] https://portal.mardi4nfdi.de/wiki/Links#Other_wikibase_instances

interdisciplinary use of resources (e.g., datasets, methods or models), FAIR metadata about resources across individual disciplines and their dependencies is crucial and lays the foundation for multi- and cross-disciplinary discovery. This is even more important given the widespread use of computational methods (e.g., NLP or ML models) across various disciplines and their dependencies on data resources. This leads to a genuine need for FAIR[13] and machine-interpretable representations of resources and their dependencies across NFDI. Some NFDI consortia already develop the knowledge graphs to support their data infrastructures, e.g., NFDI4Culture KG[14] and MaRDI Wikibase KG[15].

KGs are a key technology for the implementation of the FAIR Principles by ensuring interoperability for both machine and human access through the use of a joint data model, adoption of established and interlinked schemas and vocabularies, and reliance on consistent Persistent Identifiers (PIDs)[16] and data management policies across the entire data space.

The motivation of this working group is two-fold:

1. To promote the use of Knowledge Graphs in all NFDI consortia and facilitate cross-domain data interlinking and federation following the FAIR Principles; and
2. To contribute to the joint development of tools and technologies that enable transformation of structured and unstructured data into semantically reusable knowledge across different domains represented by NFDI consortia.

2. Objectives

- Enable transdisciplinary connection between research data assets from different consortia via an overarching KG approach.
- Assessment of the current state of KG technology implementation across the different NFDIs and identification of possible factors blocking wider adoption of KGs within individual consortia.
- Agree on an approach to support a single point of access to federated information about research assets in the NFDI, across all NFDI initiatives.

---

[13] Wilkinson, M. D., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. In Scientific Data (Vol. 3, Issue 1). Springer Science and Business Media LLC. https://doi.org/10.1038/sdata.2016.18

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

[15] https://portal.mardi4nfdi.de/wiki/Portal

[16] We acknowledge that the term 'PID' is currently used to reference a wide range of identifiers and approaches to identification in semantic contexts. We will work with other relevant NFDI WGs to come to a common understanding of the term. Furthermore, the work of relevant external communities such as the PID IG of the RDA will also be taken into consideration.

- Collaborate closely with the WG Ontology Mapping and Harmonization[17] to establish what common identifiers and connecting properties are required in order to facilitate unified access to heterogeneous resources from across the different consortia domains.
- Demonstrate how individual consortia can make their KG data (and other structured data resources) interoperable with other consortia (via federated access) in practice (cf. WG Cookbook(s), Guidance and Best Practices[18]).
- Reach out and collaborate with open source communities developing and maintaining KG technologies to ensure that particular requirements of the NFDI research infrastructure are taken into consideration and necessary steps taken towards implementing additional software customizations and extensions as required.
- Define the core requirements for a flexible and easy-to-use entry point solution for adding data to the overall knowledge graph.

3. Work Plan

Q4 2022: Evaluating the state of the art
- Define criteria and assess the current state of KG technology adoption within the NFDI consortia.
- Evaluate the use of existing integration and federation approaches (e.g., the PID Graph and its GraphQL-based implementation by DataCite[19], or cross linking via external identifiers in Wikidata).

Q4 2022 - Q4 2023: Communication activities
- Provide regular updates about WG activities.
- Knowledge sharing in NFDI working groups and consortia, and relevant external working groups.[20]
- Provide platforms for collaboration with KG tool maintainer groups (e.g. open source communities around Wikidata, Wikibase, among others) and other stakeholders.
- Prepare for and participate in NFDI Conference 2023 and Section Workshops.

Q1 2023: Overviews
- Publish an overview of KG implementations in NFDI consortia and their respective domains.

---

[17] See WG Charter: https://zenodo.org/record/6726519
[18] See WG Charter: https://zenodo.org/record/6758256
[19] This is an outcome of the FREYA H2020 Project.
[20] See the section: "Connected to other working groups outside NFDI" further in this document.

- Identify best practices and/or factors blocking the implementation of KG technology on the individual consortia-level, so that lessons learned can inform the cross-domain work as well.
- Define potential NFDI Base Services[21] based on the results of this assessment and evaluation work.

Q2 - Q3 2023: Implementation, piloting and initializing the feedback process
- Coordinate closely with the WG Ontology Mapping and Harmonization as well as NFDI Base Services teams in the implementation of a pilot cross-domain KG access point. This will involve any necessary data integration and/or schema linking activities not covered by other working groups.
- Coordinate closely with the WG Terminology Services[22] as well as Base4NFDI consortium and its respective task areas teams in the further development and improvement of KG-supporting infrastructure (including tools for LOD modeling and storage, terminology mapping, and application profile definition) in use by individual consortia.
- Coordinate testing sessions and feedback gathering with concrete cross-domain use cases for KG services.

Q3 - Q4 2023: Dissemination and documentation of outcomes
- Publication of a white paper summarizing the WG activities.
- Talks, tutorials and workshops to working groups, NFDI consortia, and other groups of interests.

Specific tasks

- Initiate regular bi-monthly and later monthly meetings for the working group.
- Evaluate existing KG tools and technologies in use within the NFDI consortia (and beyond).
- Assessment of existing approaches to achieve transdisciplinary data discovery, e.g., GraphQL.
- Identify and prioritize concrete needs for cross-domain KG access, by sub-working groups involving all consortia.
- Provide recommendations to individual consortia on the use of specific KG infrastructure setups, as well as recommendations for the overarching integrated KG approach.

---

[21] See joint statement on basic services: https://zenodo.org/record/6091657#.YgvB8t8xk2w
[22] See WG Charter: https://zenodo.org/record/6759325

- Provide example implementations of both domain-specific and cross-domain KG infrastructure setups.
- Collaborate with other relevant WGs to deliver the identifier mappings and common properties that can facilitate integrated access to a single discovery endpoint for all NFDIs.

4. Initial Membership List

	Name	Institution	NFDI consortia
1	Lozana Rossenova	TIB	NFDI4Culture
2	Markus Stocker	TIB	NFDI4DataScience, NFDI4Earth
3	Renat Shigapov	UB Mannheim	BERD@NFDI
4	Noemi Betancort	Uni Bremen	KonsortSWD
5	Felix Engel	TIB	NFDI4Ing
6	Moritz Schubotz	FIZ	MaRDI
7	Rebecca Wilm	IDS	Text+
8	Heinrich Widmann	DKRZ	NFDI4Earth
9	Claus Weiland	Senckenberg	NFDI4Earth, FAIRagro
10	Katrin Moeller	MLU Halle-Wittenberg	NFDI4Memory
11	Sonja Schimmler	Fraunhofer FOKUS	NFDI4DataScience, NFDI4Cat
12	Jakob Voß	VZG	NFDI4Objects
13	Ziyad	ZB MED	NFDI4Microbiota
14	Jürgen Kett	DNB	Text+, NFDI4Culture
15	Konrad Förstner	ZB MED	NFDI4Microbiota
16	Stefan Dietze	GESIS, HHU	NFDI4DataScience, BERD@NFDI
17	Cord Wiljes		NFDI-GS
18	Marie Annisius	DNB	Text+
19	Bridget Murphy	CAU	DAPHNE4NFDI
20	Sven Nahnsen	EKUT	GHGA DataPlant
21	Christian Bölling	MfN Berlin	NFDI4Biodiversity
22	Gerhard Heyer	SAW Leipzig	Text+
23	Thomas Eckart	SAW Leipzig	Text+
24	Daniel Mietchen	FIZ	MaRDI
25	Benjamin Zapilko	GESIS	NFDI4DataScience

5. Adoption Plan

Once assessment of the current state of KG adoption across the consortia is complete and concrete requirements for a cross-domain KG access point have been specified, feedback to the consortia and the scientific communities they represent is going to be crucial for the success of the WG.

Given the NFDI's bottom-up approach, the Section Metadata and the WG KG are the forums to discuss the best KG practices and to exchange knowledge about KGs across all NFDI domains. The WG KG is going to support the consortia in deciding which KG services and related data integration strategies should be used and how they should be modified and adopted. Furthermore, the WG will provide the expertise required in specifying the NFDI Base Services concerning KGs, and subsequently provide support to the teams developing the basic services.

Adoption plans for individual cases of KG implementation (including the cross-domain access point) may differ, depending on the state of the art within each domain. They are going to be agreed on a case-by-case basis within each consortium and with guidance from this WG as required.

Represented Consortia

NFDI4Culture

BERD@NFDI

NFDI4DataScience

NFDI4Earth

KonsortSWD

NFDI4Ing

MaRDI

Text+

NFDI4Chem

NFDI4Memory

NFDI4Cat

NFDI4Objects

NFDI4Microbiota

DAPHNE4NFDI

GHGA

DataPlant

NFDI4Biodiversity

Connected to other NFDI working groups

WG Ontology Mapping and Harmonization

WG Semantic Interoperability: Terminology Services

WG Cookbooks, Guidance, Best Practices

WG "Training Infrastructures" of Section "Training & Education"

WG "Data Integration" of Section "Common Infrastructures"

NFDI@Wikidata

Connected to other working groups outside NFDI

[Open Research Knowledge Graph](#)

[Wikimedia Germany (Wikidata)](#)

[Wikibase Stakeholders Group](#)

[Open Science Graphs for FAIR Data IG](#) (RDA)