



Facilitating creation, (re)use and interoperability of Knowledge Graphs in NFDI: perspectives from NFDI4Culture to Base4NFDI and beyond

Lozana Rossenova¹

11.07.2024

¹ TIB – Leibniz Information Centre for Science and Technology, Hannover;



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

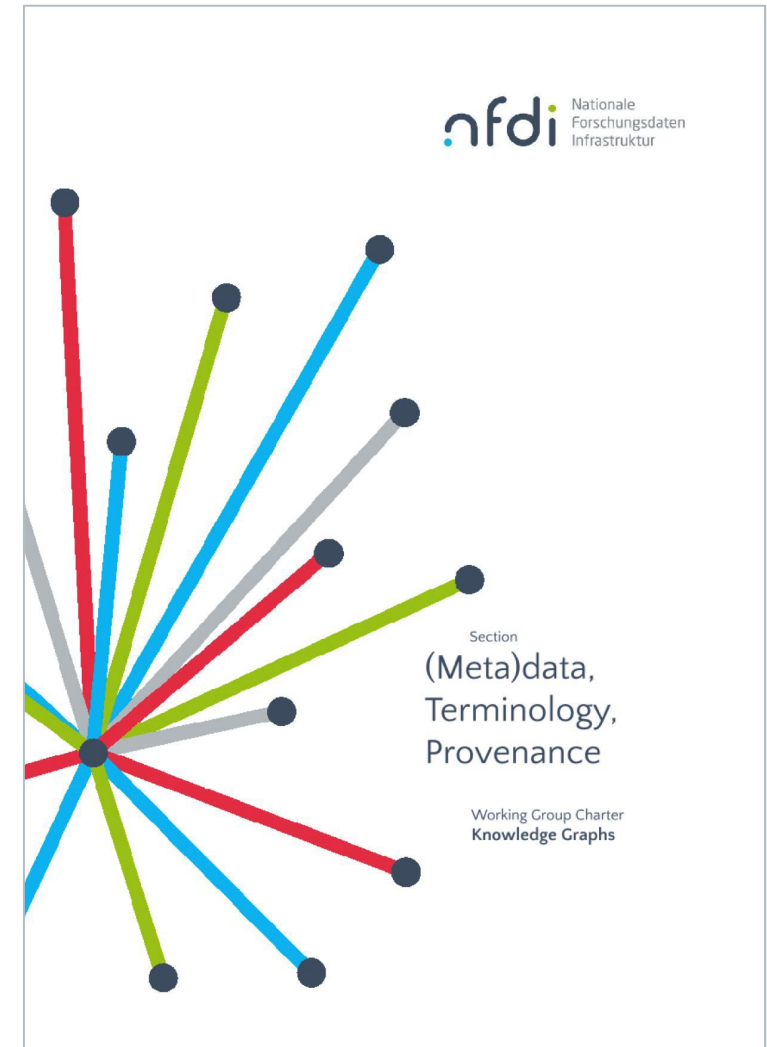
The Working Group “Knowledge Graphs” (KGs) in NFDI Section “(Meta)data, Terminologies, Provenance”

Motivation:

- Promoting the use of knowledge graphs by consortia, institutions and researchers;
- Improving FAIRness of NFDI and especially interoperability with national and international research data infrastructures;
- Contributing to development of KG tools and services.

Numbers

- 125 subscribers to the mailing list
- 56 members representing 22 consortia: the charter <https://doi.org/10.5281/zenodo.7515324>
- 3 coordinators: Renat Shigapov (BERD@NFDI), Lozana Rossenova (NFDI4Culture) & Moritz Schubotz (MaRDI)



Why KGs and why KGI?

Why KGs are an important technology for building an **interoperability framework** and enabling **data exchange**, as understood by our WG:

- KG is a **graph-structured knowledge base** containing a terminology (vocabulary or ontology) and data entities interrelated via the terminology;
- KGs are based on **semantic web technologies** (RDF, SPARQL, etc.) and often used for agile data integration;
- KGs are already **widely used** by research data producers and managers in Germany ([see poster](#));
- **Wikidata** as special connector linking between expert knowledge systems and world knowledge.

Invited talks:

1. **PID Graph & GraphQL** – Markus Stocker
2. **GESIS Search & KGI** – Benjamin Zapilko and Stefan Dietze
3. **Piveau & Data Europa** – Sonja Schimmler & Bianca Wentzel
4. **NFDI4DS Search at Uni Hamburg** – R. Usbeck, T. Taffa and A. Kraft
5. **OpenAIRE Research Graph** – Andreas Czerniak

Why KGs and why KGI?

Humanities and social sciences

- BERD@NFDI (KGs)
- KonsortSWD
- NFDI4Culture (KGs)
- NFDI4Memory (KGs)
- NFDI4Objects
- Text+ (KG)

Engineering sciences

- NFDI4DataScience (KGs & KG Software)
- NFDI4Energy (KG)
- NFDI4Ing (KG Software)
- NFDI-MatWerk (KGs & KG Software)
- NFDIxCS

Life sciences

- DataPLANT
- FAIRagro
- NFDI4Immuno
- GHGA
- NFDI4Biodiversity
- NFDI4BIOIMAGE (KG)
- NFDI4Health
- NFDI4Microbiota (KG)

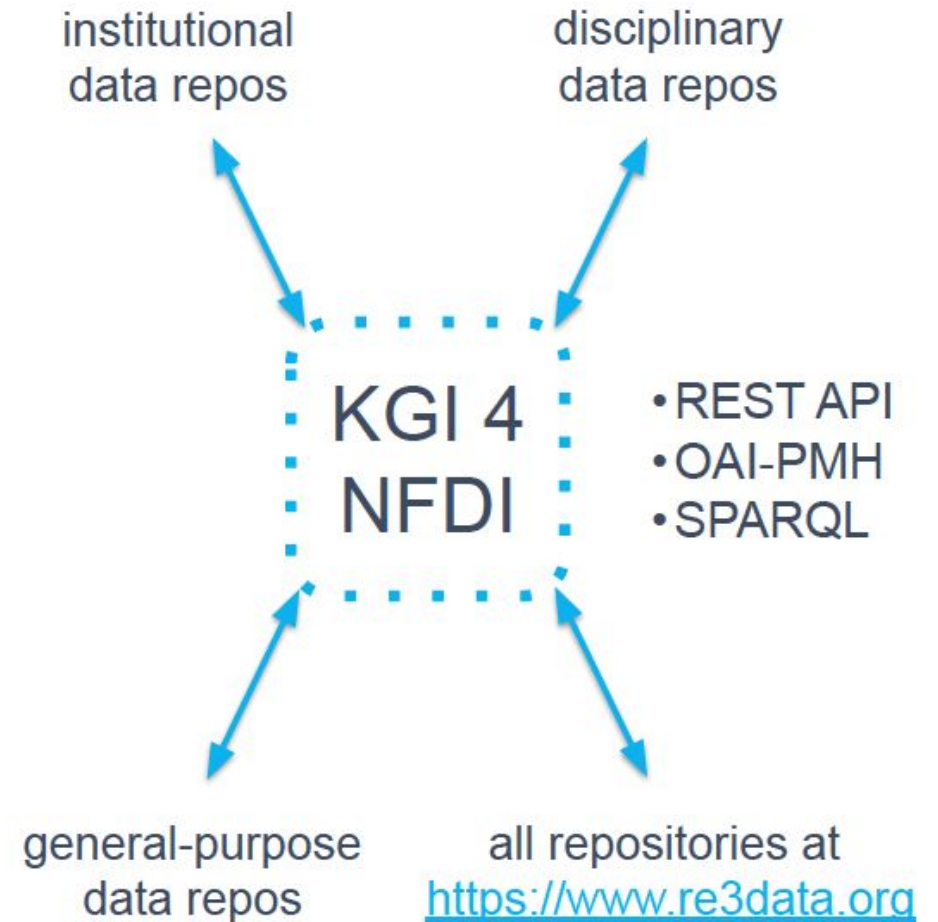
Natural sciences

- DAPHNE4NFDI
- FAIRmat
- NFDI4Cat (KG)
- MaRDI (KGs)
- NFDI4Chem (KGs)
- NFDI4Earth (KG)
- PUNCH4NFDI

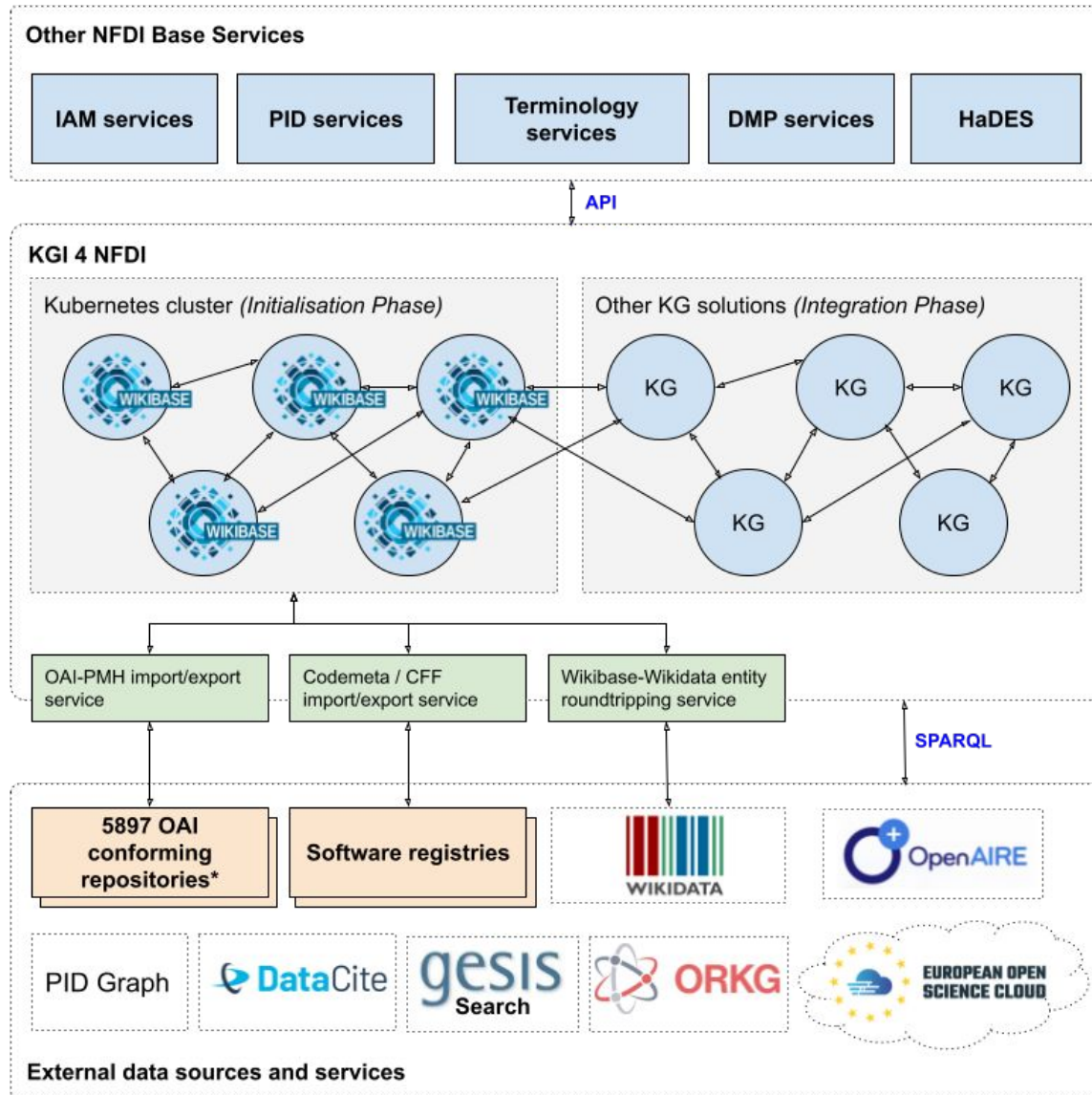
Why KGs and why KGI?

NFDI needs to be **interoperable internally and across national and international research data infrastructures** (as the section WGs testify):

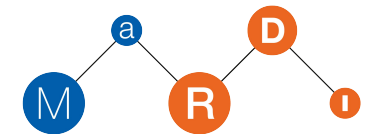
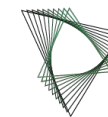
- Individual solutions may be required to meet **domain-specific** requirements;
- NFDI needs an **interoperable network** of metadata knowledge graphs (RDF, SPARQL);
- Consortia, institutions and researchers need an easy-to-use, scalable and interoperable **KGI-as-a-Service**.



KGI-as-a-Service proposal 1.0



...an ecosystem of software, including tools for data import, validation and export, collaborative frontends, search APIs and SPARQL endpoints with result visualization, Extract-Transform-Load and data linking software...



* Source: <https://www.openarchives.org/Register/BrowseSites>

Original proposal to Base4NFDI

Proposal submitted on 15.02.2023:

- Combining the **ease-of-use** of software like Wikidata with research-backed data;
- Allowing NFDI stakeholders to **create KGs** without administrative overhead;
- Developing an **interoperability framework** for connecting KGs with research infrastructures;
- And establishing a **KGI-consultancy** to increase adoption of the KGI-service.

Pilot phase based on one specific tool suite as a **'minimum viable product'** (Wikibase):

- Landscape analysis (learning the needs of consortia and researchers; overview papers);
- Deployment scalability (Kubernetes cluster);
- Interoperability pipelines (OAI-PMH & Codemeta / CFF import/export to SPARQL);
- Consultancy (help with creating knowledge graphs).

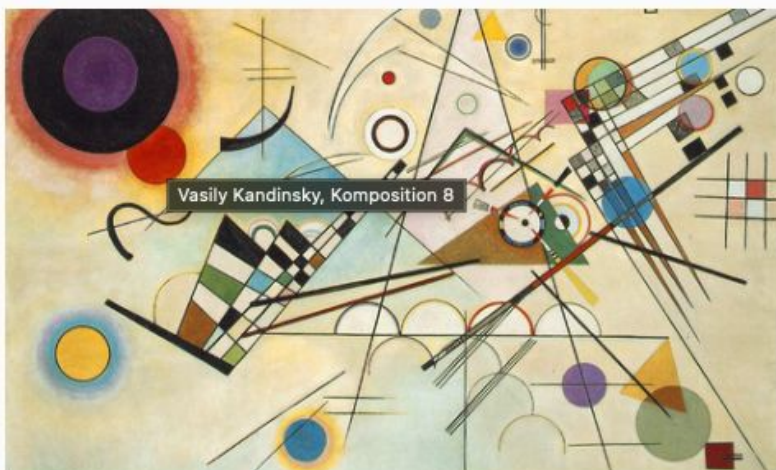
Choice of pilot software suite and use cases

Wikibase and Wikidata adoption:

- Wikidata KG, already used by various consortia and participating institutions – both as a **repository** to upload data to, and a rich resource on the linked open data (LOD) cloud to **federate** with;
- **Growing adoption** of Wikibase and the popularity of Wikidata as proof-of-concept;
- Mix of human- and machine-readable interfaces can lower the barrier to **participation**.

Use cases:

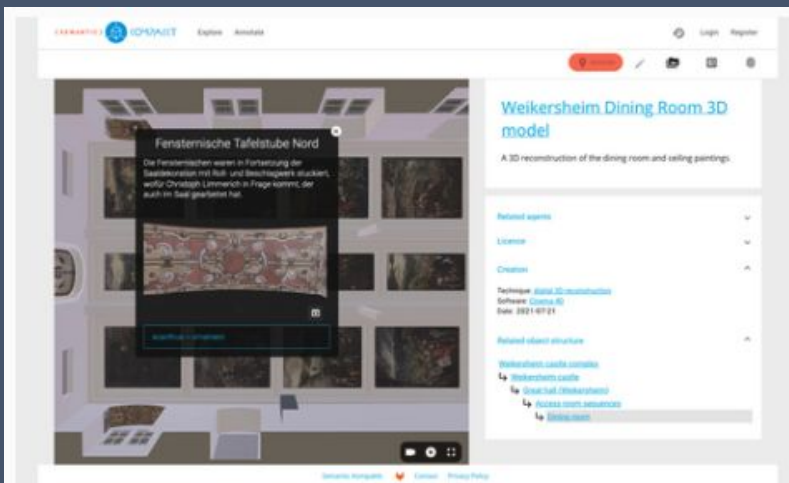
- **MaRDI** and **BERD4NFDI** are using Wikibase instances as central portals for all research data;
- **NFDI4Culture** offer Wikibase instances to annotate digitized cultural objects with structured data;
- **NFDI4Memory** includes FactGrid, Wikibase instance hosted at the University of Erfurt, as central repository for data about historical persons and events.



Culture Knowledge Graph

The Culture Knowledge Graph aims to be a connector for all research data produced within the NFDI4Culture subject areas.

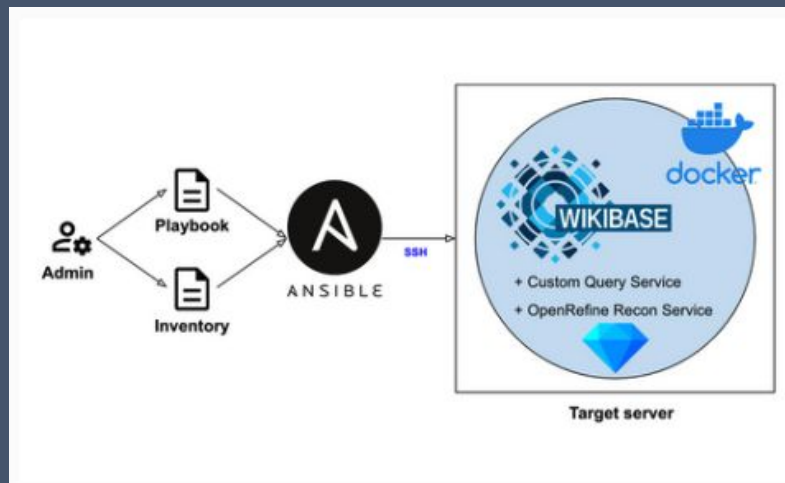
[Find out more](#)



Semantic Kompakkt

Semantic Kompakkt is a free and open source toolchain for the viewing and annotation of 3D models, and other visual media within a linked open data (LOD) environment.

[Find out more](#)



Wikibase4Research

Wikibase4Research is a free and open source suite of tools for the storage and management of Linked Open Data (LOD).

[Find out more](#)

← → ↻ https://wikibase.semantic-kompakt.de/wiki/Main_Page

CnFdI
3D DATA
ENRICHMENT

Hauptseite Diskussion

Main Page

Inhaltsverzeichnis [Verbergen]

- 1 Semantic annotation for 3D cultural artefacts
- 2 About the original case study
- 3 Adding new data in the archive
- 4 Example item pages for different types of data
- 5 Data model reference
- 6 Some example data queries
 - 6.1 Federated Queries
- 7 Indexes for quick reference

Semantic annotation for 3D cultural artefacts

A suite of tools for semantic annotation of 3D cultural artefacts is being developed (see [the Open Science lab at TIB, Hannover](#)). Operating within Task area 1: Digital Cultural Heritage within a knowledge graph environment, so that 3D objects' geometry, attributes and metadata are not lost. The project builds on several existing FOSS tools:

- [OpenRefine](#), a data cleaning, reconciliation and batch upload tool;
- [Wikibase](#) (the tool behind the interface you are viewing now), a suite of tools for semantic annotation of digital culture

```
14 #Query wikidata
15 SERVICE wdgqs: {
16
17 #Find castles with renaissance architectural style
18 ?castle wdt:P31 wd:Q751876.
19 ?castle wdt:P149 wd:Q236122.
20
21 #Look for those castles in a radius of 100km around our castle
22 SERVICE wikibase:around {
23   ?castle wdt:P625 ?location .
24   bd:serviceParam wikibase:center ?coordinates .
25   bd:serviceParam wikibase:radius "100" .
26 }
27
28 #Get labels from Wikidata
29 ?castle rdfs:label ?castleLabel.
30 OPTIONAL { ?castle wdt:P18 ?image. }
31 FILTER((LANG(?castleLabel)) = "de")
32 }
33 }
```

Map 8 results in 2769 ms <> Code Download Link

Point(9.89583 49.4806)
Schloss Weikersheim

NFDI4Memory: FactGrid

https://database.factgrid.de/wiki/Main_Page

Welcome!


on the FactGrid database, a project of the Gotha Research Centre operated by the data lab of the University of Erfurt. With the support of Wikimedia Germany we are using a MediaWiki with Wikidata's "wikibase" extension. Our main product are data which we collect on "items" such as these:

- Adam Weishaupt
- Paris

You will need to be logged in to see the "add statement" link with which you can add information in the form of triple based machine readable claims — which everyone can now explore with the SPARQL data mining language, at our Query Service. All FactGrid data are CC0-licensed. You can download any search in various data formats with the aim to explore FactGrid data in other software environments or visualise searches with various tools on our site.

Visit our

- FactGrid FAQ for more information on why you might love to use FactGrid as your research platform
- Help Section for further assistance



https://database.factgrid.de/query/#%23defaultView%3ATimeline%0ASELECT%3FReid



1750	1760	1770	1780	1790	1800	1810	1820
			●	●			
			12. Februar 1784				
			Adolph Freiherr Knigge zu Gast bei Christoph Bode, Arbeit an der gemeinsamen Erklärung, Weimar, 1784-02-12				
			11. Februar 1784				
			Adolph Freiherr Knigge zu Gast bei Christoph Bode's, Gespräche über Freimaurerei und Illuminaten, Weimar, 1784-02-11				
			16. Juli 1782				
			Wilhelmsbader Konvent, 15. Juli bis 1. Sept. 1782				
			1. Mai 1776				
			Gründung des Perfectibilisten-Ordens (1778 in Illuminatenorden umbenannt)				
			13. Februar 1784				
			Freiherr Knigge zu Gast bei Christoph Bode, Gedanken über einen Nachfolger der Strikten Observanz, Weimar, 1784-02-13				
			13. Februar 1784				
			Knigge und Bode Abends zu Gast bei Goethe, Über den Ursprung des Illuminatenordens, Weimar 1784-02-13				

Outcome and feedback on first proposal

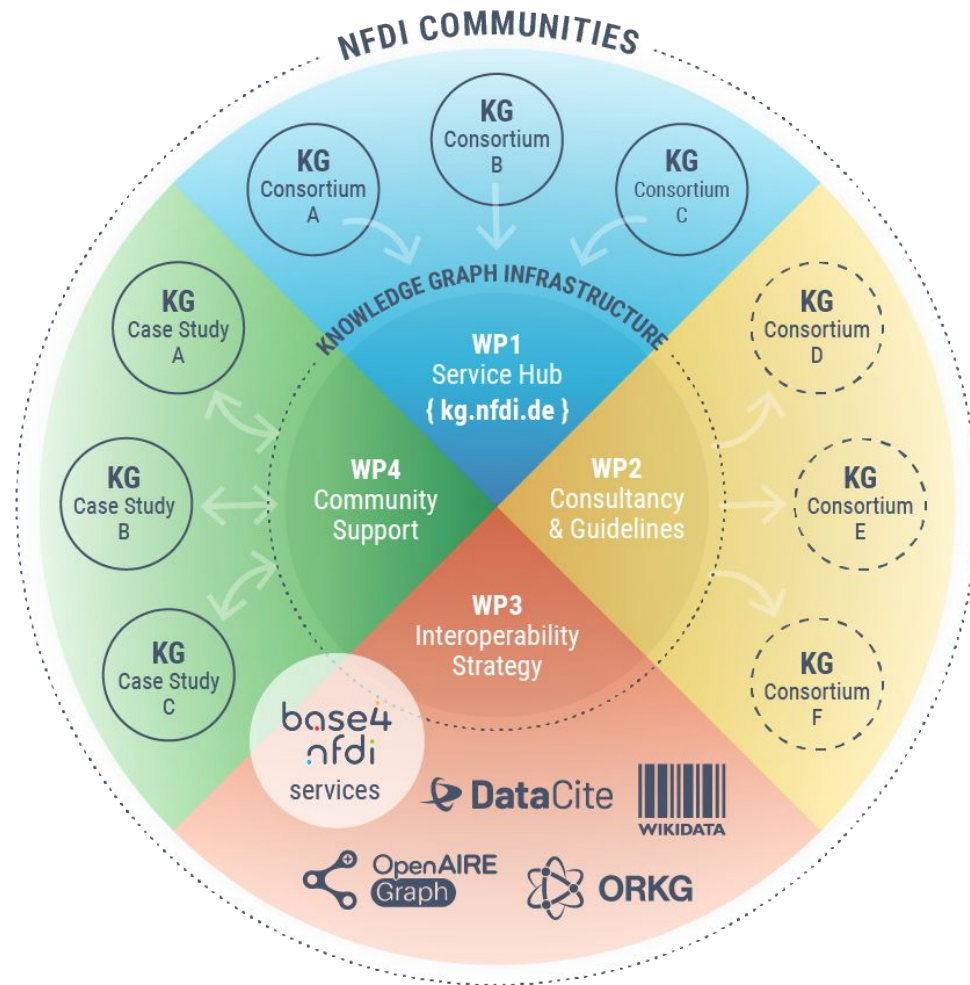
Unsuccessful as basic services, suggested changes:

- Include **use cases** from more consortia;
- Better explain how the **different software solutions** already in place can be integrated;
- Gather support from more consortia at **voting** stage (especially important for later funding phases);

Lessons learned:

1. Natural and life sciences have **other data workflows**, not accounted for in case studies we considered for pilot phase.
2. **Ontology and terminology service** questions need to be solved independently from concrete KG infrastructure solutions.
3. **Service-orientation** of Base4NFDI doesn't provision for implementation of one specific open source solution.

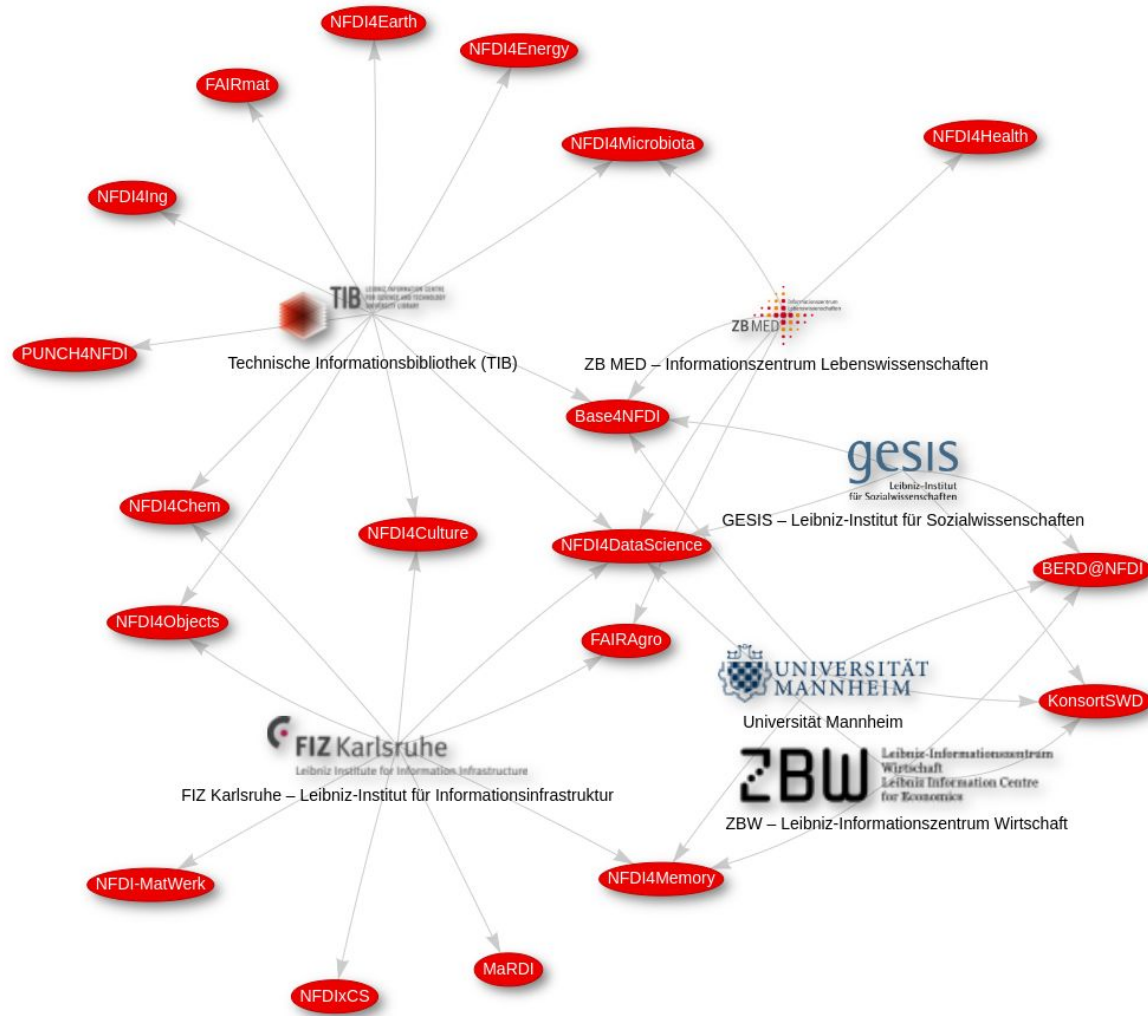
KGI-as-a-Service proposal 2.0



Legend: ○ Existing KGs ○ KGs in the making

...will include a KG registry, which will aggregate information on all KGs contributed by NFDI consortia and by the research communities they represent, as well as a service to facilitate access to KGs across NFDI projects. Furthermore, the base service aims to empower research communities to create KGs by providing the necessary technologies and expertise for decentralised KG instances based on standards and tried-and-tested approaches....

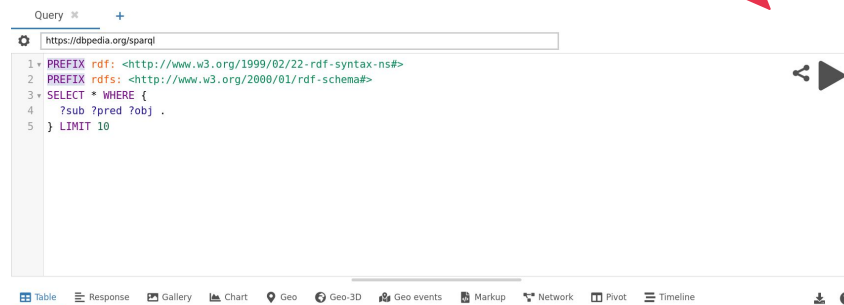
KGI-as-a-Service proposal 2.0: team



- FIZ Karlsruhe (MaRDI)
- GESIS (NFDI4DataScience)
- TIB (NFDI4Culture)
- Uni Mannheim Library (BERD@NFDI)
- ZB MED (NFDI4Microbiota)
- ZBW (KonsortSWD)

Work Package 1 - Central services

- D1.1 Central Registry of Knowledge Graphs in NFDI
 - registry in the form of a knowledge graph to host metadata about NFDI services that rely on KGs;
 - adoption of a simple ontology (based on ongoing work in several consortia);
 - editorial and curatorial process via common, easy-to-contribute-to and collaborative tool (e.g. GitHub, or Wikidata);
 - RDF-based ELT workflow (based on NFDI4Culture) on a timed schedule;
- D1.2 Platform for search and query across KGs
 - central website (Flask web framework and additional Python libraries);
 - SPARQL editor and preprocessor;
 - example queries.



```
Query x +
https://dbpedia.org/sparql
1 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
2 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
3 SELECT * WHERE {
4 ?sub ?pred ?obj .
5 } LIMIT 10
```



NFDI consortium	Knowledge Graph	URL
BERD@NFDI	Aktienführer KG	https://akf.kgi.uni-mannheim.de
BERD@NFDI	MaschinenBauIndustrie KG	https://mbi.kgi.uni-mannheim.de
MaRDI	MaRDI portal	https://portal.mardi4nfdi.de
NFDI4Cat	4Cat Meta Portal	https://meta4cat.fokus.fraunhofer.de/ (/sparql)
NFDI4Culture	Culture Knowledge Graph	https://nfdi4culture.de/resources/knowledge-graph
NFDI4Culture	Linked Stage Graph	https://slod.fiz-karlsruhe.de/about
NFDI4Culture	Semantic Kompakkt	https://wikibase.semantic-kompakkt.de
NFDI4DataScience	Open Research Knowledge Graph	https://orkg.org
NFDI4DataScience	TweetsKB	https://data.gesis.org/tweetskb
NFDI4DataScience	TweetsCov19	https://data.gesis.org/tweetscov19
NFDI4DataScience	ClaimsKG	https://data.gesis.org/claimskg
NFDI4DataScience	SoftwareKG	https://data.gesis.org/softwarekg
NFDI4DataScience	SoMeSci	https://data.gesis.org/somesci
NFDI4DataScience	Question Feature Sample	https://data.gesis.org/questionfeaturesample
NFDI4DataScience	4DS Meta Portal	https://meta4ds.fokus.fraunhofer.de/ (/sparql)
NFDI4DataScience	dblp computer science bibliography	https://sparql.dblp.org/dblp
NFDI4Earth	NFDI4Earth KG	https://nfdi4earth-knowledgehub.geo.tu-dresden.de/fuseki/dataset.html
NFDI4Ing	SciMesh	https://scimesh.org/about
NFDI4Memory	FactGRID	https://database.factgrid.de
NFDI4Microbiota	InteractOA (using Wikidata)	https://interactoa.toolforge.org
Text+	Gemeinsame Normdatei (GND)	https://www.dnb.de/EN/Professionell/Metadaten/endienste/Datenbezug/LDS/lids_node

Work Package 2 - Consultancy and Guidelines

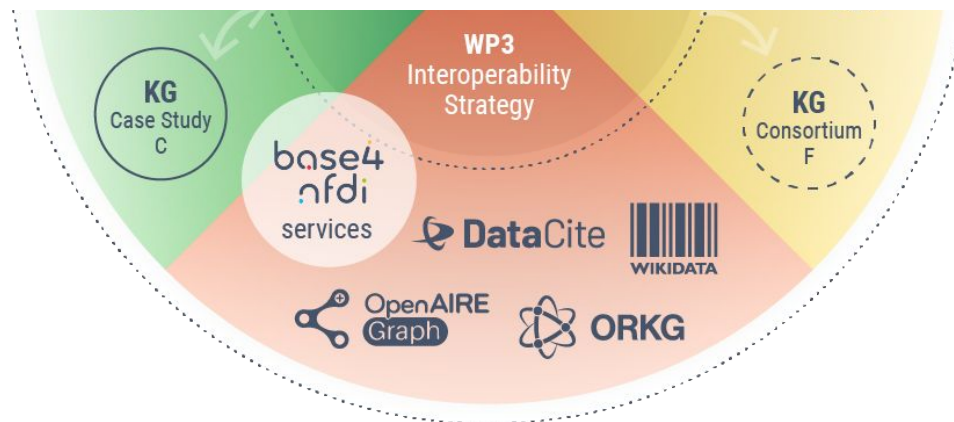
- D2.1 Guidelines for creating and hosting Knowledge Graphs
 - set of guidelines, step-by-step instructions, educational videos for:
 1. Installation & Configuration
 2. Data Modeling & Import
 3. Querying
 - Ansible playbooks for deployment;



- D2.2 Consultancy service
 - consultancy service with central contact point available via the main service hub;
 - regular office hours as well as 1-to-1 consulting sessions on demand.

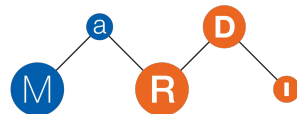
Work Package 3 - Interoperability Strategy

- D3.1 Strategy for metadata mapping, linking, and integration
 - based on:
 - existing and well-established standards & vocabularies – schema.org, Dublin Core, NFDIcore ontology, etc;;
 - APIs for data sharing, such as OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting);
 - PIDs for consistent identification across all data;
 - strategy document which supports necessary consensus-building processes by providing guidelines and best practices;
- D3.2 Strategy for interoperability with national and international KG initiatives
 - strategy to establish interoperability with: OpenAIRE Graph, PID Graph, ORKG, Helmholtz KG and Wikidata;
 - identification of basis for potential collaboration and joint activities.



Work Package 4 - Community Support

- D4.1 Survey for the existing KG adoption practices in NFDI
 - two objectives:
 - (a) identify requirements and feedback (benefits and challenges) relating to KG adoption; and
 - (b) monitor, to the extent possible, KG adoption across the NFDI, including new use cases, technology preferences/solutions, and so on.
 - collaboration with Base4NFDI team and service steward;
 - compare survey findings with international initiatives (EOSC, RDA) and monitor changes to the state of the art.
- D4.2 Show cases
 - case studies to monitor and assess how the KGI basic service impacts consortia, and what synergies it offers;
 - incubator process following the example of the NFDI AAI basic service to grow number over time;
 - monitor development and/or adoption across consortia and facilitate exchange between KGI and the selected case studies;
 - select the “query of the month” to show the value of connected NFDI KGs;
 - examples of ontology harmonisation for KGs from different consortia with close topical proximity.

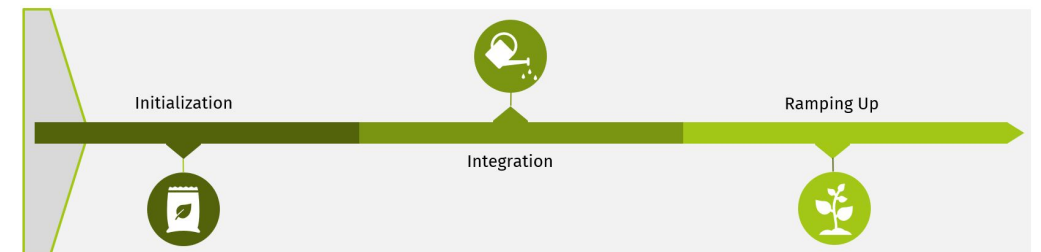


Plans for the Integration Phase

- Integration of a Natural Language Interfaces using LLMs;
- Tighten integration of / interaction with other Base4NFDI services;
- Transitioning experimental KGs into stable stages;
- Supporting ontology harmonization and mapping (according to strategy from D3.1);
- Extension of collection of Show Cases;
- Interaction with EOSC services;

Plans for the Ramp-up Phase

- Increase service performance and increase responsiveness;
- Increase reliability and monitoring;
- Reduce issue resolution time.



Outlook and next steps

- Kick off event for KGI team in July 2024;
- Setting up public mailing list;
- Hiring staff;
- Survey preparation and completion within first quarter of the service initialization;
- Testing and evaluating software for the KGI service hub.