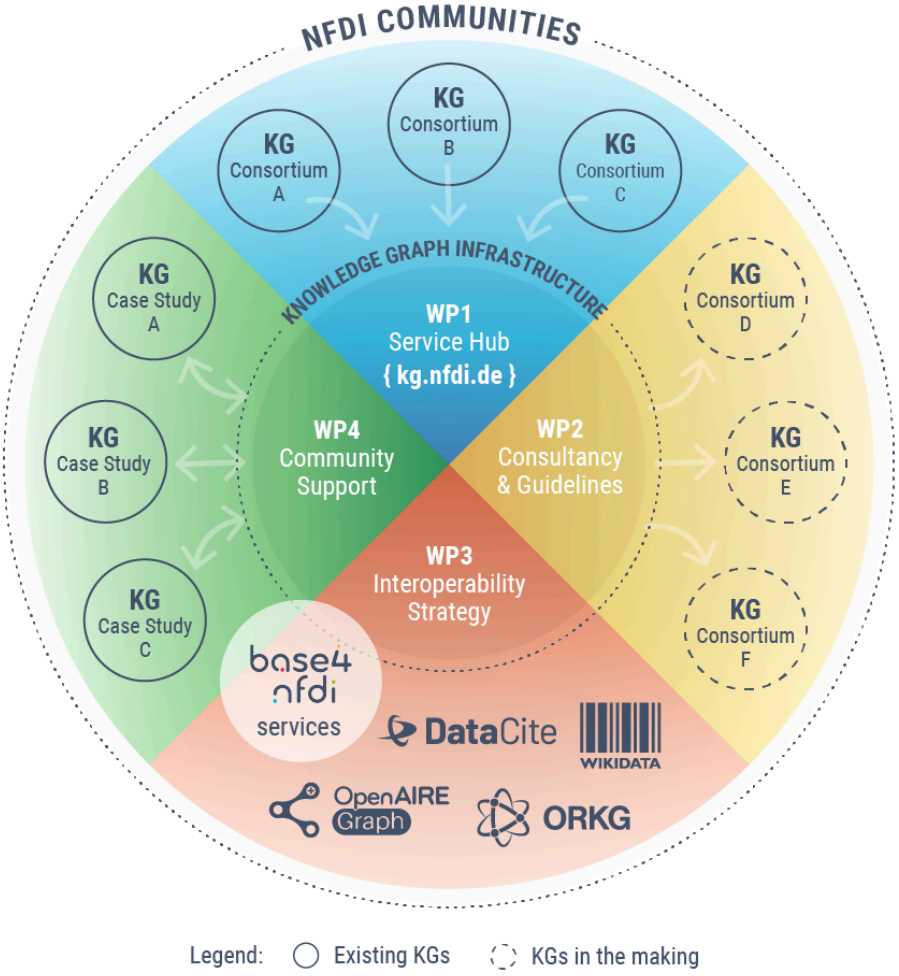
KGI4NFDI

Knowledge Graph Infrastructure for the German National Research Data Infrastructure



Proposal for the Initialisation Phase of Base4NFDI

Submitted: February 15th, 2024

On behalf of: WG Knowledge Graphs, Section Metadata, Terminologies, Provenance

https://doi.org/10.5281/zenodo.13118749

I Proposal - Initialisation Phase

1. General Information

Knowledge Graph Infrastructure (KGI4NFDI)

A service hub to help consortia, institutions and researchers create and link knowledge graphs to ensure interoperability within the NFDI and across international research data infrastructures.

Lead institution

ZB MED – Information Centre for Life Sciences
Gleueler Strasse 60
50931 Cologne

Name of lead institution principle investigator

Prof. Dr. Konrad Förstner (foerstner@zbmed.de)

Participating institutions

Principal Investigator	Institution, location	Contact E-mail	Member in [consortium]
Prof. Dr. Konrad Förstner	ZB MED – Information Centre for Life Sciences, Cologne	foerstner@zbmed.de	NFDI4Microbiota
Dr. Lozana Rossenova	TIB – Leibniz Information Centre for Science and Technology, Hannover	lozana.rossenova@tib.eu	NFDI4Culture
Dr. Renat Shigapov	Mannheim University Library, Mannheim	renat.shigapov@uni-mannheim.de	BERD@NFDI
Dr. Moritz Schubotz	FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, Berlin	moritz.schubotz@fiz-karlsruhe.de	MaRDI
Fidan Limani	ZBW – Leibniz Information Centre for Economics, Kiel	flimani@zbw-workspace.eu	KonsortSWD
Dr. Benjamin Zapilko	GESIS - Leibniz Institute for the	Benjamin.Zapilko@gesis.org	NFDI4DataScience

	Social Sciences, Mannheim		

Table 1: List of participating institutions

Partner institutions without funding

Institution, location	Input	Contact person, email	Member in [consortium]
Wikimedia Deutschland – Gesellschaft zur Förderung Freien Wissens e. V., Berlin	Developers of Wikibase, Wikibase.Cloud and Wikidata	Dr. Raja Amelung, raja.amelung@wikimedia.de	NFDI4DataScience
Fraunhofer FOKUS, TU Berlin	Expertise on research knowledge graphs, meta portals	Prof. Dr. Sonja Schimmler, sonja.schimmler@fokus.fraunhofer.de	NFDI4DataScience, NFDI4Cat
Forschungszentrum Jülich GmbH & Helmholtz Metadata Collaboration (HMC)	Input for Harmonisation Strategy re. Helmholtz KG	Dr. Volker Hofmann, v.hofmann@fz-juelich.de	NFDIMatWerk
Heidelberg Institute for Theoretical Studies (HITS)	Show case for the medical field	Martin Golebiewski, martin.golebiewski@h-its.org	NFDI4Health
Helmholtz Centre for Ocean Research (GEOMAR), Kiel & Helmholtz Metadata Collaboration (HMC)	Input for Harmonisation Strategy re. Helmholtz	Sören Lorenz, slorenz@geomar.de	-
OpenAIRE	Input for Harmonisation Strategy considering the OpenAIRE Graph	Paolo Manghi, paolo.manghi@openaire.eu	-

Table 2: List of partner institutions that will contribute to this proposal without receiving funding. Partially they are planning to be involved further in the next phases.

Planned duration of the project: 1 year

Summary of the proposal in English

Knowledge graphs (KG) are an important and powerful tool for achieving interoperability in the research domain and fulfilling the NFDI's mission. While several consortia are already building their own KG solutions, what is currently missing is a central and reusable KG infrastructure (KGI). The proposed base service will provide key components of such a KGI. These 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. To refine and optimise these services, a comprehensive survey will be conducted. Documentation on the efficient creation of KGs will be compiled, and the base service will also offer dedicated consulting services to consortia and facilitate harmonisation of ontologies across different consortia. Overall, this integrated approach will seek to enhance data integration and foster standardised ontological practices. It will make a significant contribution to advancing the “One NFDI” vision and will support the FAIRification of data across disciplines. The base service will complement international KG initiatives, such as those promoted by the European Open Science Cloud (EOSC).

Summary of the proposal in German

Wissensgraphen (Knowledge graphs, KG) sind ein wichtiges und leistungsfähiges Werkzeug für Interoperabilität im Forschungsbereich und zur Erfüllung der Mission der NFDI. Während bereits mehrere Konsortien ihre eigenen KG-Lösungen entwickeln, fehlt eine zentrale und wiederverwendbare KG-Infrastruktur (KGI). Dieser Basisdienst wird Komponenten einer solchen KGI bereitstellen. Dazu gehören ein KG-Registry, das Informationen zu allen von NFDI-Konsortien und ihrer Forschungsgemeinschaften beigetragenen KGs aggregiert, sowie ein Dienst zur Erleichterung des Zugangs zu diesen KG. Zudem zielt der Basisdienst darauf ab, Forschungsgemeinschaften zu ermächtigen, KGs zu erstellen, indem er die nötige Technologien und Fachkenntnisse für dezentrale KG-Instanzen auf Basis von Standards und erprobten Ansätzen bereitstellt. Zur Verbesserung dieser Dienste wird eine umfassende Umfrage durchgeführt. Es wird Dokumentation zur Erstellung von KGs zusammegetragen, und der Basisservice wird dedizierte Beratungsdienste für Konsortien anbieten und die Harmonisierung von Ontologien verschiedener Konsortien erleichtern. Dieser Ansatz wird darauf abzielen, die Datenintegration zu verbessern und standardisierte ontologische Praktiken zu fördern. Er wird einen Beitrag zur Förderung der Vision von "One NFDI" leisten und die FAIRifizierung von Daten über Disziplinen hinweg unterstützen. Der Basisservice wird internationale KG-Initiativen wie diejenigen des European Open Science Cloud (EOSC) ergänzen.

2. State of the Art of Proposed Basic Service

2.1 Background and Motivation

A *knowledge graph* (KG) is a graph-structured knowledge base containing a terminology (vocabulary or ontology) and data entities interrelated via the terminology (Hogan, *et al, 2022*). Knowledge graphs are based on semantic web technologies[1] such as RDF

[1] https://www.w3.org/standards/semanticweb

(Resource Description Framework) and SPARQL (SPARQL Protocol And RDF Query Language) and are often used for agile data integration. The *Knowledge Graph Infrastructure* (KGI) described in this proposal envisions an entire ecosystem of software dedicated to the creation of KGs, including tools for data import, validation and export, collaborative frontends, search APIs, SPARQL endpoints, and tools for visualising query results and KG registries. One prominent example of an existing KGI is the Wikibase technology stack with its vast collection of tools and its widely used instance, Wikidata. The adoption of KGs in science is a significant step towards making data more FAIR (findable, accessible, interoperable, and reusable). By leveraging KGs, researchers can create structured and interconnected representations of diverse datasets, thereby fostering a holistic understanding of human knowledge. The Strategic Research and Innovation Agenda (SRIA) of the European Open Science Cloud (EOSC) identifies Knowledge Graphs (KG) as one of the most important technologies for building an interoperability framework and enabling data exchange among users across countries, sectors, and disciplines (European Commission, 2022). KGs play an essential role within the German research community as a vehicle to facilitate data exchange between institutions and individual researchers.

Key goals of KGI4NFDI: The "Knowledge Graph Infrastructure" (KGI4NFDI) project seeks to deliver core knowledge-graph services to the National Research Data Infrastructure (NFDI) consortia and their associated research communities. To do this, it will create a **registry of knowledge graphs used by NFDI consortia**, which will itself be presented in the form of a knowledge graph. In addition, the project will provide **guidance and documentation** on the creation and maintenance of knowledge graphs, devise an **interoperability strategy**, conduct a **survey**, and **showcase** the application of knowledge graphs in diverse research scenarios across various fields. The implementation will build upon a commonly used **FLOSS (Free, Libre Open Source Software) technology stack** and the generated solutions will themselves be made available under open source and content licences to maximise reusability and sustainability.

2.2 State of the art

The use of KGs in science and the humanities facilitates the seamless integration and linking of data, thereby promoting collaboration and interdisciplinary research. Wikidata exemplifies the application of knowledge graphs by providing a centralised hub of structured data which is used by researchers to integrate and link information across various domains, enhancing discoverability and accessibility. Another example is the Open Research Knowledge Graph (ORKG), a platform dedicated to scholarly knowledge. Some long-standing resources have now also been made available in the form of KGs[2], such as UniProt, a comprehensive

[2] https://sparql.uniprot.org/

database for protein information. Moreover, initiatives such as UnHIDE (Unified Helmholtz Information and Data Exchange)[3] strive to interconnect data sources across various disciplines within a significant scientific organisation.

Currently, several NFDI consortia are in the process of building KG solutions or are already providing KG-compatible resources. However, in several consortia, the technical and organisational overhead to establish and run these services can be hard to justify, as our interview with representatives of the consortia revealed (see below). A further disadvantage is the lack of a multidisciplinary registry of KGs and their features. This makes it difficult to run efficient queries across multiple KGs and makes KGs less discoverable.

2.3 Status of work results in preparation for the Basic Service

The Working Group "Knowledge Graphs" (WG KGs) was established in the NFDI Section "Metadata, Terminologies, Provenance" to coordinate the development and use of KGs in all NFDI consortia. The WG has since carried out an evaluation of the state of the art of KG adoption in the NFDI. The primary use cases of KGs in the NFDI were identified as (1) data integration across repositories, institutions, consortia and the NFDI; and (2) access to and control of (meta)data via SPARQL endpoints. The institutions involved in this base service are already co-applicants and/or participants in 17 of the 27 NFDI consortia.[4] Additionally, interviews with representatives of all 26 thematic consortia were conducted at the end of 2023 and the beginning of 2024[5]. These interviews were used to identify the overarching needs that are addressed in this proposal. All consortia were invited to comment on the proposal before its submission. In the interviews, several consortia stated that they had already established KGs, while some only have initial small-scale prototypes, and others have not yet started using KGs at all. Of the consortia that have not yet used KGs, the majority intend to use them in the future or already have concrete plans to do so. When asked about the desired support that a base service could offer, the consortia that did not yet have an established KG expressed the need for technical support, including on-site consulting and ready-to-use software solutions that they could implement themselves. Plans to connect different KGs within their own consortium (and also to link them to the KGs of other consortia) were mainly mentioned by the consortia that already have existing KGs, but also by some whose KGs are still at the planning stage. Several consortia also expressed the need for a central hub with a KG registry. While the knowledge graphs compiled in such a registry will be technically interoperable, they might lack semantic interoperability (i.e. harmonised ontologies). Nonetheless, an up-to-date registry of KGs and their ontologies

[3] https://docs.unhide.helmholtz-metadaten.de / https://search.unhide.helmholtz-metadaten.de/
[4] https://w.wiki/99YJ (Wikidata query with resulting graph) - see also Appendix
[5] https://zenodo.org/doi/10.5281/zenodo.10654409

would, for example, be hugely beneficial for the work of the “Ontology Harmonization and Mapping” working group and would lay the technical groundwork for the required and crucial negotiation processes. Some interviewees also requested training courses on querying and visualising KGs as well as opportunities to share expertise and experiences. Key criteria for KGI4NFDI's choice of suitable KG technologies included technological flexibility and the ability to offer multiple options. A useful starting point was the experience gained by NFDI4Culture: a survey carried out by this consortium revealed low overall adoption of KGs within the cultural heritage domain. To address the challenges of heterogeneous and distributed cultural research data, NFDI4Culture implemented a KG-based RDM system with an accessible web interface[6], as well as an RDF-based automated ELT and workflow management system (Sack, et al, 2024). This work provides an excellent basis for the KGI base service to build upon and could potentially be adopted by other consortia.

2.4 Current Technical Readiness Level (TRL) of the proposed Basic Service

The envisioned service will build upon several different technologies (see Working Packages for further details). Apache Jena (TBD2 and Fuseki2) stands out as a highly mature technology, achieving a TRL of 9. Similarly, Virtuoso (TRL 9) is known for its performance and versatility in managing large-scale RDF triple stores. Wikibase (TRL 9) has been extensively adopted by organisations worldwide as a base for structured data compilations with proven scalability (e.g. Wikidata). YasGUI (TRL 9) provides an intuitive interface for querying triple store instances and enhances user accessibility. WebVowl (TRL 9) offers powerful visualisation capabilities, facilitating the exploration and understanding of complex ontologies. Wikibase.cloud, with a TRL of 8, represents a slightly lower but still substantial level of maturity. This cloud-based variant of Wikibase extends the platform's capabilities to a distributed environment. It is provided by Wikimedia Deutschland, which confirmed that the service will be maintained and extended in the foreseeable future (see Letter of Intent). Ansible represents a dependable deployment tool with a TRL of 9.

3. SWOT Analysis

A SWOT analysis of the pilot KGI service is provided below.

Internal	**Strengths** 1. The project team leads or is active in several consortia covering a broad range of topics and has a strong track record in implementing KGs for research data management.	**Weaknesses** 1. Many different needs must be met in relation to KGs. *Mitigation*: Showcase examples of KG use and foster long-term engagement and support strategies to include all stakeholders.

[6] https://nfdi4culture.de/services/details/culture-knowledge-graph.html

	2. The proposal was created based on the input of all 26 NFDI consortia and taking different needs into account. 3. The project will provide significant support to the work of other cross-cutting Section Working Groups e.g. Knowledge Graphs and Ontology Mapping and Harmonisation. 4. Primarily based on TRL 8 and 9 software that has been proven in an operational environment (e.g. Wikibase solutions with more than 100 million entities, such as Wikidata). 5. Offers a holistic approach that integrates solutions for technical and operational implementation.	2. Not all members of the NFDI community are familiar with the benefits that KGs provide or the established standards for their creation and use; they are thus unable to give informed feedback (in the survey). *Mitigation: Use showcases to present benefits and guidelines and provide consulting services to encourage informed community engagement with the KG base solutions.*
External	**Opportunities** 1. Knowledge graphs are technically compliant with the FAIR principles and offer the opportunity for semantic interoperability. They are suitable tools for forging connections across research fields and for implementing the “One NFDI” vision. 2. Ability to connect NFDI data to existing KG-based services that promote open research and machine-actionable data enrichment. 3. Establish strategies for streamlining access to data – from the local consortia level to the international level (e.g. EOSC, OpenAIRE).	**Threats** 1. Dependency on information provision by all NFDI consortia. *Mitigation: Ensure proactive involvement of all consortia (as in the preparation of this proposal) and identify knowledgeable representatives that can be included in the negotiation processes.* 2. Other technical approaches for consortium-level data integration become favoured over KGs. *Mitigation: Data integration is only one of the use cases for KGs. The KGI service can still be used for many other use cases, e.g., for providing access to data via SPARQL endpoints or query federation.* 3. Lack of consensus among consortia about ontology mapping principles would reduce interoperability of the KGI with other NFDI components. *Mitigation: All NFDI requirements will be collected in the initialisation phase and will be taken into account in the integration phase. Close collaboration with Section Metadata and its related Working Groups will aim to support*

		consensus-seeking in situations where lack thereof may cause issues for the KGI service.

Table 2: SWOT Analysis

4. Working Concept for the Development of the Basic Service

The working concept for the KGI service is structured around **two primary goals**: **1.** Provide a service hub where research communities can find, explore and query KGs; **2.** Support research communities in the efficient creation of KGs by providing tools and consulting services, and assist with the harmonisation of ontologies across research fields.

To achieve these goals, the following **five objectives** have been identified: **1.** Build a service hub with a KG registry; **2.** Build capacity for KG concepts and applications; **3.** Showcases; **4.** Enable queries across research fields (harmonisation and interoperability); **5.** Enable easier interaction with KGs e.g. by providing natural language interfaces.

4.1 Service initialisation concept

In the **initialisation phase**, we will establish the service hub using pilot knowledge graphs from selected NFDI consortia and provide support in the creation of new KGs and the expansion of existing ones. This will be carried out through four work packages (WPs).

1. Establish the service hub: An entry point to all KGs provided by the different NFDI consortia will be established at kg.nfdi.de (or a similarly prominent location). In addition to serving as the gateway for accessing KGs from different consortia, this hub will facilitate seamless interactions with KGs across disciplines while showcasing the “One NFDI” vision. A comprehensive KG registry will offer detailed insights into the structure and content of each KG. A sustainable, collaborative curatorial process for data entry and editing of the registry will be developed and disseminated. The registry itself will also be implemented as a KG that can be queried via SPARQL. It will also include a curated library of example queries spanning diverse domains.

2. Provide consulting and documentation: This WP will provide advice on building KGs while also gathering feedback on what support is required in order to grow KG adoption across all consortia. We will produce and publish the required tutorials and documentation open access. The consulting team will provide expertise in at least three key software solutions already widely used by NFDI consortia (Wikibase/Wikidata, Virtuoso and Apache Jena).

3. Facilitate interoperability: This WP aims to ensure interoperability with existing standards, best practices, and activities inside and outside of NFDI. The output will be a strategy for metadata mapping, linking, and integration, both conceptually (on, e.g. a schema

and vocabulary level) and technically (through the inclusion of established standard tools and services). Priority will be given to requirements specifications (e.g. the survey carried out in WP4) and applied technologies (e.g. the NFDIcore ontology (Sack *et al.*, 2024)) from other consortia. Furthermore, a strategy will be developed for the interaction with existing KGs, such as OpenAIRE Graph, PID Graph, ORKG, Helmholtz KG and Wikidata. These activities will actively support the process of building consensus on interoperability standards and the derivation of best practices.

4. Strengthen community: At the end of the initialisation phase, the requirements of all relevant consortia, institutions and researchers for KGI services will be clarified and synthesised into detailed proposals for the next phases. We hope to use the pilot KGI as a test case for how other solution platforms could be effectively deployed and made interoperable within the overall KGI architecture. By working with software solutions with high TRLs, we hope to move through the initialisation phase without significant hurdles and with sufficient time to extract best practices that can be applied in later phases.

4.2 Development and integration outlook

The development and integration phase will focus on the transition to one connected network of KGs. It will also work on enhancing the user interface and experience, for example via the experimental introduction of Natural Language Interfaces (NLIs) using Large Language Models (LLMs). This phase will also see a concerted effort to integrate deeply with other Base4NFDI services (TS4NFID, PID4NFDI and IAM4NFDI) in order to foster a more interconnected and seamless ecosystem. A pivotal task will be to transition experimental KGs into stable, reliable stages, ensuring they are ready for broader use. Additionally, we will continue supporting ontology harmonisation and mapping (see details in D3.1). We will also expand our collection of showcases (WP4). A further task will be to enhance our interaction with other services, particularly the European Open Science Cloud (EOSC) and OpenAIRE Graph services.

4.3 Ramping up for Operation

As we transition into the ramping-up-for-operation phase, our focus shifts towards optimising the performance and responsiveness of our services. This entails scaling up our infrastructure to not only handle increased demand, but also to ensure that our systems are operating at their optimal capacity. Enhancing reliability and implementing comprehensive monitoring mechanisms are critical steps towards achieving a high level of service availability and performance. A significant goal during this phase is to minimise issue resolution times and reach TLR-9 for all services. However, a lively community is necessary and, to achieve that, we need to ensure that the ecosystem stays healthy and leaves room

for new initiatives.

4.4 Risks and challenges

This section outlines the general risks and challenges not covered in the SWOT analysis table, addressing the three phases of the project (initialisation, integration, and ramping-up-for-operation) and the selection criteria[7] provided by Base4NFDI.

1. Interoperability in NFDI: This can become a challenge if interoperability with certain NFDI services is not considered in the initialisation phase. *Risk-mitigation measures: We will integrate with the existing basic services IAM (WP1), PID (WP3) and TS (WP3). Moreover, our case studies and survey (WP4) will incorporate ongoing feedback from the consortia.*

2. International interoperability: Though the KGI service aims to achieve international interoperability (particularly in regard to EOSC), there is a risk that certain national and international infrastructures will not be taken into account in the initialisation phase. *Risk-mitigation measures: WP3 will develop a dedicated strategy for interoperability, which will include international interoperability.*

3. Limited scalability: Large KGs such as Wikibase become difficult to manage when their size reaches a certain limit. The initialisation phase does not include plans to make queries across KGs more efficient. *Risk-mitigation measures: Based on the slow-query logs generated, we will suggest optimisation strategies for the next phases.*

4. Recruiting qualified staff: Difficulty in recruiting highly specialised staff within the timeframe of the WPs and schedule of deliverables. KG technologies require niche expertise that is hard to find. *Risk-mitigation measures: The partnering organisations already include staff that can be involved in the project.*

[7] https://base4nfdi.de/how-base4nfdi-will-decide

5. Work Programme

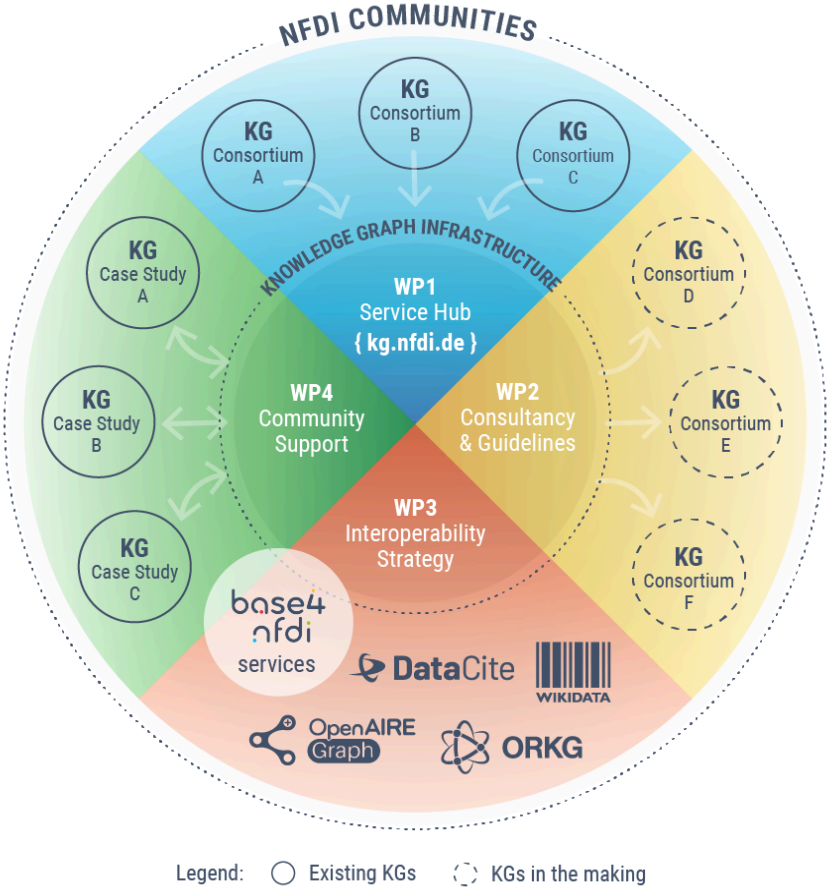


Figure 1: Overview of the work programme.

5.1. Overview of work packages

Work package	Deliverables (D) and Milestones (M)	Responsible partner(s)
1. Service Hub	D1.1 Registry of KGs in NFDI D1.2 Platform for search and query across KGs M1.1 Curatorial process guidelines M1.2 Service hub established	ZB MED / TIB
2. Consultancy and Guidelines	D2.1 Guidelines for creating and hosting KGs D2.2 Consultancy service M2.1 Guidelines available online M2.2 Consultancy service established	UB Mannheim / TIB
3. Interoperability Strategy	D3.1 Strategy for metadata mapping, linking, and integration D3.2 Strategy for interoperability with national and international KG initiatives M3.1 Release of strategy documentation	GESIS / ZBW
4. Community Support	D4.1 Survey of existing KG adoption practices in NFDI D4.2 Showcases M4.1 Release of survey results M4.2 Monthly activity reports that will be announced via a newsletter	FIZ / ZBW

Table 3: Overall work programme with work packages, deliverables, milestones & responsible partner(s).

5.2. Detailed work programme

Work Packages, Tasks, Deliverables and Milestones	Base Service "Knowledge Graph Infrastructure"											
	1	2	3	4	5	6	7	8	9	10	11	12
WP1: Service Hub						M1.1						M1.2
Task / Deliverable D1.1						D1.1						
Task / Deliverable D1.2												D1.2
WP2: Consultancy and Guidelines						M2.1						M2.2
Task / Deliverable D2.1						D2.1						
Task / Deliverable D2.2												D2.2
WP3: Interoperability												M3.1
Task / Deliverable D3.1						D3.1						
Task / Deliverable D3.2												D3.2
WP4: Community				M4.1								M4.2
Task / Deliverable D4.1				D4.1								
Task / Deliverable D4.2												D4.2

Chart 1: Gantt chart with work packages **(WPx)**, deliverables **(Dx.y)**, and milestones **(Mx.y)**.

5.2.1 WP1: Service hub (months 1-12)

In this work package, the central technical services will be established and maintained. A central registry of knowledge graphs hosted and/or used by the different NFDI consortia will be created. The registry itself will be implemented as an openly available and editable knowledge graph that can be queried via a SPARQL endpoint. A central website will display and explore the registered knowledge graphs and offer visual tools and dashboards to access information about them.

Deliverable D1.1 Registry of knowledge graphs in NFDI: After formalising the results of the interview-based requirements analysis, we will create a registry in the form of a knowledge graph to host metadata about NFDI services that rely on KGs. This will include the adoption of a simple ontology to describe this metadata collection based on ongoing work in several consortia (Sack, *et al.*, 2023). Instead of starting from scratch, we will build upon existing efforts to collect metadata on KGs in use and their endpoints created within the NFDI[8] and beyond[9], thereby consolidating existing initiatives rather than creating yet another siloed resource. Thanks to active participation in their disciplinary fields, national and international infrastructure initiatives and relevant cross-cutting NFDI WGs, the project

[8] https://zenodo.org/records/8124286
[9] https://www.wikidata.org/wiki/Wikidata:Lists/SPARQL_endpoints; https://www.re3data.org/; https://dataportals.org/

partners are well positioned to coordinate such activities. Crucially, we will develop an editorial and curatorial process that will be coordinated with the relevant WGs in order to guarantee long-term sustainability of our approach. In consultation with the relevant WG stakeholders and active representatives of the consortia, we will select a common, easy-to-contribute-to and collaborative tool (e.g. GitHub, or Wikidata), with version history, that will act as a gathering point and empower stakeholders to keep their own information up to date. Guidelines will be established (in coordination with WP2) to safeguard quality, accuracy, and consistency of the overall data contribution process. We will use Apache Jena TBD2 as a triple store and Apache Jena Fuseki2 as a SPARQL endpoint[10] to create a KG with the data gathered via the collaborative curatorial process, as this represents a performant, scalable and widely used open source technology stack. Data will be periodically ingested into the KG using an RDF-based ELT (Extract, load, transform) workflow established for NFDI4Culture (Sack *et al.)* on a timed schedule. A central website will be created using the Flask web framework[11] and additional Python libraries. For the visualisation of ontologies, a WebVOWL instance will be hosted and embedded into the web platform[12]. Further data, such as the number of entities, will be visualised using a dashboard solution created for NFDI4Culture. The data compiled will be available under a CC0 licence and will also be regularly deposited in relevant open repositories, including an RDF dump on Zenodo.

Milestone: *M1.1 Curatorial process guidelines created and disseminated to stakeholders (M6)*

Deliverable D1.2 Platform for search and query across KGs: To make search and query across KGs possible, a SPARQL editor and preprocessor, Yasgui[13], will be embedded into the web platform. For all knowledge graphs registered, example queries will be provided that show the possibilities of RDF-based data management. Additionally, more-complex example queries that span different knowledge graphs and research fields will be included. Those examples will initially be created as part of discussions between the participating consortia in close collaboration with work package WP4, which will also verify that the queries run in different environments. The queries will be compiled in a dedicated git repository on Github, enabling the community to suggest improvements and submit their own example queries.

Milestone: *M1.2 Service hub established (M12)*

---

[10] https://jena.apache.org/
[11] https://flask.palletsprojects.com
[12] https://github.com/VisualDataWeb/WebVOWL
[13] https://docs.triply.cc/yasgui/

5.2.2 WP2: Consultancy and Guidelines (months 1-12)

WP2 will establish communication channels between the basic service and NFDI consortia, Sections and Working Groups on the topic of KGs. Consulting sessions will be organised to ensure effective knowledge transfer. The feedback gathered during these sessions will be crucial in tailoring the guidelines documentation to specific user needs. The required guidelines will be created and openly published under a Creative Commons Attribution licence (CC-BY).

Deliverable D2.1 Guidelines for knowledge graph creation and hosting: We will develop and compile a set of guidelines that provide step-by-step instructions, educational videos and best practices for creating and hosting knowledge graphs. These guidelines will be specific to at least three key software solutions already referenced by NFDI communities during the preliminary interviews[14]: Wikibase[15] (self-hosted or via Wikibase.cloud[16]), Virtuoso[17], and Apache Jena[18]. These solutions are already used by several consortia, and the possibility remains of expanding this offer to include additional solutions if requested by research communities surveyed in WP4. The general structure of these guidelines will be organised into three main sections: 1) “Installation & Configuration” will provide instructions on installing and configuring software. 2) “Data Modeling & Import” will explain data modelling & data import functionalities. 3) “Querying” will describe SPARQL querying functionalities.

***Milestone**: M2.1 Guidelines available online (M6)*

Deliverable D2.2 Consultancy service: We will establish a consultancy service with a central contact point available via the main service hub (kg.nfdi.de). The service will offer regular office hours as well as 1-to-1 consulting sessions on demand. A dedicated KGI consultant will onboard users and service providers in the guidelines (D2.1) and in the central registry and query platform (D1.1, D1.2). Consulting sessions will also seek to gather further feedback from the research communities and translate this into service requirements to be actioned out in the initialisation phase or later phases of the project, where appropriate.

***Milestone**: M2.2 Consultancy service established (M12)*

5.2.3 WP3: Interoperability Strategy (months 1-12)

The objective of WP3 is to ensure interoperability between the KGs of different consortia as well as with national / international KG initiatives. Thus, in this WP we will focus on the

[14] https://doi.org/10.5281/zenodo.10654410
[15] https://wikiba.se
[16] https://www.wikibase.cloud
[17] https://virtuoso.openlinksw.com
[18] https://jena.apache.org

development of strategy documents for (a) metadata mapping, linking, and integration among KGs, and (b) interoperability with national and international KG initiatives. This WP lays the foundations for the negotiation processes facilitating semantic interoperability across research fields.

Deliverable D3.1 Strategy for metadata mapping, linking, and integration: We will elaborate a strategy for metadata mapping, linking, and integration for KGs. For this strategy, we will consider existing and well-established W3C standards, e.g., shared vocabularies such as schema.org, Dublin Core, and DataCite; APIs for data sharing, such as OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting); PIDs for consistent identification across all data; and others. The survey results from WP4, in particular requirements and currently used tools and standards, will be considered. Relevant developments already undertaken in other consortia, e.g., the NFDIcore ontology (Sack et al, 2024), will also be taken into account. The outcome of D3.1 will be a strategy document which supports necessary consensus-building processes by providing derived guidelines and best practices for KG metadata mapping, linking, and integration, including overviews of currently used tools and standards.

D3.2 Strategy for interoperability with other significant KG initiatives: Several KG initiatives and activities can be observed at a national and international level, including, for example, OpenAIRE Graph, PID Graph, ORKG, Helmholtz KG and Wikidata. In this deliverable, we will elaborate a strategy to establish interoperability with these initiatives. Starting with an overview of existing initiatives and activities, including the technologies and standards used, we will identify intersections and common areas of interest which could form the basis for potential collaboration and joint activities. The outcome of D3.2 will be a strategy document for establishing exchange and interoperability with other KG initiatives, including the areas and activities identified as ripe for collaboration.

Milestone: *M3.1 Release of strategy documentation (M12)*

5.2.4 WP4: Community support (months 1-12)

Establishing and maintaining communication with the community is essential to driving adoption of the KGI service. To achieve this, we propose two measures to encourage uptake, each of which is specified as a deliverable: D4.1 “Showcases” will evaluate the overall effect of the KGI activities based on selected examples, and D4.2 “Survey of existing KG adoption practices in NFDI” will complement this by systematically monitoring the quality of the basic service.

Deliverable D4.1 Survey of existing KG adoption practices in NFDI: A survey is a suitable tool for assessing the number of KGs adopted across all NFDI projects, for eliciting

user requirements, and for collecting feedback from the potentially large number of NFDI members who are involved, directly or indirectly, in KG adoption. We aim to reach at least 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. The survey will be conducted in close collaboration with the Base4NFDI team and the service steward responsible for semantic interoperability & reusability provided by Base4NFDI. Collecting requirements has already started. The Knowledge Graphs Working Group (KG WG)[19] has collected information on KG adoption across NFDI projects[20] and carried out a series of interviews with NFDI members to understand current KG adoption across NFDI projects. Based on the results of this initial work, we plan to conduct one survey across the NFDI projects within the first quarter of the project lifecycle. This allows us to also include NFDI members not covered by the KG WG process so far, while leaving enough time for feedback analysis. The survey results will be incorporated and reflected in relevant project tasks and deliverables. Given that KG adoption is also gaining momentum outside the NFDI community, it is important to examine how other initiatives are dealing with KG adoption. Thus, we also plan to compare the project survey findings with those from international initiatives, such as the EOSC[21] and RDA[22], in order to monitor any changes to the state of the art in KG adoption.

Milestone: *M4.1 Release of survey results (M4)*

Deliverable D4.2 Showcases: We will prepare several case studies, including various NFDI projects such as NFDI4Microbiota, NFDI4Culture, NFDI4DataScience, BERD@NFDI, MaRDI and NFDI4Health. These will enable us to monitor and assess how the KGI basic service impacts consortia, and what synergies it offers. The case studies will start out as a small set and will increase in number over time, with an incubator process following the example of the NFDI AAI basic service. We will closely monitor the development and/or adoption of the technology in the consortia and facilitate the exchange between KGI and the selected case studies. The case-study coordinator will cooperate with designated contacts from the participating consortia. Together with the partner consortia, we will select the “query of the month” to show the value of connected NFDI KGs. We will also provide examples of ontology harmonisation for KGs from different consortia with close topical proximity and explore the required negotiation processes which will generate relevant experiences to contribute to the interoperability strategy (D3.1.).

Milestones*: M4.2 Monthly activity reports that will be announced via a newsletter*

[19] https://zenodo.org/records/7515324.
[20] See https://zenodo.org/records/8332776
[21] https://eosc.eu/
[22] https://www.rd-alliance.org/groups/open-science-graphs-fair-data-ig

6. Funding Request

[...]

7. Required Support Actions from Base4NFDI / NFDI Sections / NFDI consortia

Support from	Work package	Contact Person Basic Service
Base4NFDI; Section Common Infrastructures; Section Metadata	**WP 1 - Service hub:** Support in coordinating compliance with other related Base Services, such as IAM (Identity and Access Management), TS, PIDs, etc.	Konrad Förstner; Sonja Schimmler
Base4NFDI; Section Metadata WG “Knowledge Graphs”	**WP 2 - Consultancy:** Support in promoting the consultancy offering;	Renat Shigapov; Lozana Rossenova
Base4NFDI; Section Common Infrastructures; Section Metadata WG “Ontology mapping and harmonization”	**WP 3 - Interoperability strategy:** Support in coordinating compliance with other related Base Services, as well as related interoperability activities in individual consortia;	Benjamin Zapilko
Base4NFDI; Section Metadata WG “Knowledge Graphs”	**WP 4 - Community support:** Support from Service Steward in conducting the survey	Fidan Limani; Moritz Schubotz

Table 8: Support request

III Appendix

a) Bibliography and list of references

Joseph Corneli, and **Moritz Schubotz.** 2017. math.wikipedia.org: A vision for a collaborative semi-formal, language independent math (s) encyclopedia. *2nd Conference on Artificial Intelligence and Theorem Proving.* URL: http://aitp-conference.org/2017/aitp17-proceedings.pdf

Dennis Diefenbach, Max De Wilde, and Samantha Alipio. 2021. Wikibase as an Infrastructure for Knowledge Graphs: The EU Knowledge Graph. In: *The Semantic Web – ISWC 2021. ISWC 2021. Lecture Notes in Computer Science*, vol 12922. Springer, Cham. https://doi.org/10.1007/978-3-030-88361-4_37

Muhammad Elhossary, **Konrad U. Förstner**, InteractOA: Showcasing the representation of knowledge from scientific literature in Wikidata, Semantic Web journal, *under review*, https://semantic-web-journal.net/content/interactoa-showcasing-representation-knowledge-scientific-literature-wikidata-0

European Commission, Directorate-General for Research and Innovation. 2022. Strategic Research and Innovation Agenda (SRIA) of the European Open Science Cloud (EOSC). Publications Office of the European Union. URL: https://data.europa.eu/doi/10.2777/935288

Aidan Hogan, Claudio Gutierrez, Michael Cochcz, Gerard de Melo, Sabrina Kirranc, Axel Pollcrcs, et al. 2022. *Knowledge Graphs.* Springer Cham. URL: https://doi.org/10.1007/978-3-031-01918-0

Daniel Mietchen, Gregor Hagedorn, Egon Willighagen, et al. 2015. Enabling Open Science: Wikidata for Research (Wiki4R). *Research Ideas and Outcomes 1:* e7573. doi: https://doi.org/10.3897/rio.1.e7573.

Robert Nasarek, Lozana Rossenova, Lucia Sohmen, and Paul Duchesne. 2023. Forschungsdaten-Management-Services für Linked Open Data: Ein Vergleich. *DHd 2023 Open Humanities Open Culture, 13–17 March 2023, Trier, Germany / Belval, Luxembourg and online.*

Finn Årup Nielsen, **Daniel Mietchen**, and Egon Willighagen. 2017. Scholia and scientometrics with Wikidata. *Joint Proceedings of the 1st International Workshop on Scientometrics and 1st International Workshop on Enabling Decentralised Scholarly Communication. CEUR Workshop Proceedings. 1878.* URL: https://doi.org/10.5281/zenodo.1036595.

André Greiner-Petter, **Moritz Schubotz**, et al. 2022. Do the math: Making mathematics in Wikipedia computable. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* URL: https://doi.org/10.1109/TPAMI.2022.3195261

Lozana Rossenova, Paul Duchesne, and Ina Blümel. 2022. Wikidata and Wikibase as complementary research data management services for cultural heritage data. *In: Proceedings of the 3rd Wikidata Workshop 2022, co-located with the 21st International*

Semantic Web Conference (ISWC2022), Virtual Event, Hangzhou, China, October 2022. URL: https://ceur-ws.org/Vol-3262/paper15.pdf

Harald Sack, Torsten Schrade, Oleksandra Bruns, Etienne Posthumus, Tabea Tietz, Ebrahim Norouzi, Jörg Waitelonis, Heike Fliegl, Linnaea Söhn, Julia Tolksdorf, Jonatan Jalle Steller, Abril Azócar Guzmán, Said Fathalla, Ahmad Zainul Ihsan, Volker Hofmann, Stefan Sandfeld, Felix Fritzen, Amir Laadhar, Sonja Schimmler, and Peter Mutschke. (2024, January 31). Knowledge Graph-Based Research Data Management - NFDI4Culture, NFDI-MatWerk, NFDI4DataScience and more. NFDI4Memory Ontologies Workshop, Bayerische Staatsbibliothek, München. Zenodo. https://doi.org/10.5281/zenodo.10600183

Renat Shigapov, and Irene Schumm. 2021. BERD: The knowledge graph of German companies. *Wikibase in Knowledge Graph based Research Data Management (NFDI) Projects.* URL: https://madoc.bib.uni-mannheim.de/58793

Renat Shigapov. 2022. Knowledge graphs in BERD and in NFDI. *Focused Tutorial on Capturing, Enriching, Disseminating Research Data Objects. Use Cases from Text+, NFDI4Culture and BERD@NFDI, Mannheim and online.* Zenodo. URL: https://doi.org/10.5281/zenodo.7373258.

Markus, Stocker, **Lozana Rossenova, Renat Shigapov**, Noemi Betancort, Stefan Dietze, Bridget Murphy, Christian Bölling, **Moritz Schubotz**, and Oliver Koepler. 2023. Knowledge Graphs – Working Group Charter (NFDI Section-Metadata) (1.1). Zenodo. URL: https://doi.org/10.5281/zenodo.7515324

Andra Waagmeester, Gregory Stupp, Sebastian Burgstaller-Muehlbacher, Benjamin M Good, Malachi Griffith, Obi L Griffith, Kristina Hanspers, Henning Hermjakob, Toby S Hudson, Kevin Hybiske, Sarah M Keating, Magnus Manske, Michael Mayers, Daniel Mietchen, Elvira Mitraka, Alexander R Pico, Timothy Putman, Anders Riutta, Nuria Queralt-Rosinach, Lynn M Schriml, Thomas Shafee, Denise Slenter, Ralf Stephan, Katherine Thornton, Ginger Tsueng, Roger Tu, Sabah Ul-Hasan, Egon Willighagen, Chunlei Wu, and Andrew I Su. 2020. Wikidata as a knowledge graph for the life sciences. *eLife 2020*; 9:e52614. DOI: 10.7554/eLife.52614.

Wikimedia. 2021. Strategy 2021: Wikibase ecosystem. URL: https://meta.wikimedia.org/wiki/LinkedOpenData/Strategy2021/Wikibase.

b) Graph visualisation of the participating institutions and their NFDI consortia affiliations

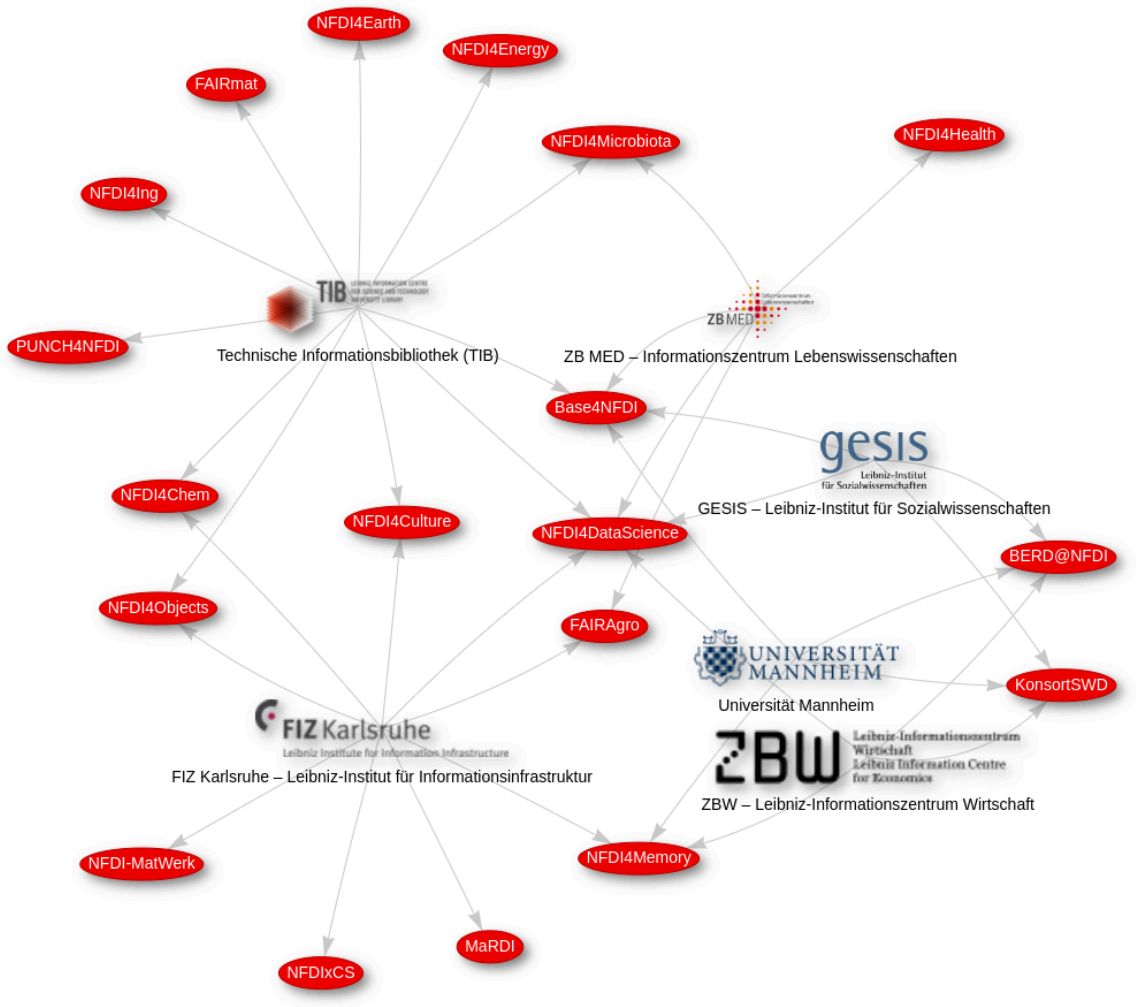


Figure 2: Co-applicant and participant affiliations to NFDI consortia. Data sourced from Wikidata: https://w.wiki/99YJ

c) Letters of support by the consortia supporting the basic service development