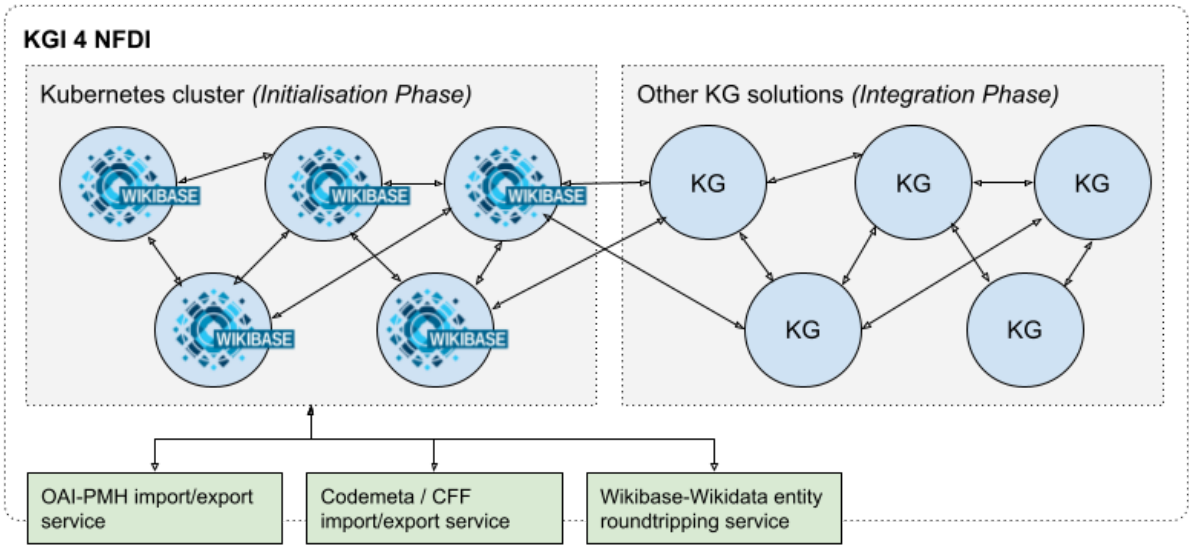
KGI4NFDI

Knowledge Graph Infrastructure for the German National Research Data Infrastructure



Proposal for the Initialization Phase of Base4NFDI

Authors: Lozana Rossenova[1], Renat Shigapov[2], Moritz Schubotz[3] and Daniel Mietchen[3]

Submitted: February 15th, 2023

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

1 TIB – Leibniz Information Centre for Science and Technology, Hannover;
2 University Library | University of Mannheim, Mannheim
3 FIZ Karlsruhe - Leibniz Institute for Information Infrastructure, Berlin

I Proposal - Initialization Phase

1. General Information

- Name of proposed Basic Service (in English)

  Knowledge Graph Infrastructure

- Acronym of the proposed Basic Service

  KGI

- Service "subtitle" explaining key functionality

A service for deploying knowledge graphs by consortia, institutions and researchers to ensure interoperability within NFDI and across international research data infrastructures

- Lead institution

  TIB – Leibniz Information Centre for Science and Technology

  Lange Laube 28

  30159 Hannover

- Name of lead institution principle investigator

  Dr. Lozana Rossenova

- Participating institutions

Principal Investigator	Institution, location	Member in [consortium]
Dr. Lozana Rossenova	TIB – Leibniz Information Centre for Science and Technology, Hannover	NFDI4Culture
Dr. Renat Shigapov	Mannheim University Library, Mannheim	BERD@NFDI
Dr. Moritz Schubotz	FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, Berlin	MaRDI

Table 1: List of participating institutions

- Partner institutions

Institution, location	Knowledge graph expertise	Contact person	Member in [consortium]
Wikimedia Deutschland – Gesellschaft zur Förderung Freien Wissens e. V., Berlin	Developers of Wikibase, Wikibase.Cloud and Wikidata	Dr. Raja Amelung	NFDI4DataScience
ZB MED - Information Centre for Life Sciences, Cologne	Applications of Wikidata/Wikibase and ORKG in the Life Sciences	Prof. Dr. Konrad Förstner	NFDI4Health, NFDI4Microbiota, NFDI4DataScience, FAIRAgro
GESIS - Leibniz Institute for the Social Sciences	Construction, hosting and provisioning of research knowledge graphs for Social Sciences, AI & Computer Science	Prof. Dr. Stefan Dietze	KonsortSWD, BERD@NFDI, NFDI4DataScience, Base4NFDI
Fraunhofer FOKUS, TU Berlin	Research knowledge graphs, meta portals	Dr. Sonja Schimmler	NFDI4DataScience, NFDI4Cat, Base4NFDI, Sektion Infra
ORKG – Open Research Knowledge Graph, TIB, Hannover	Knowledge-Graph-based representation and augmentation for scientific research publications	Prof. Dr. Sören Auer	NFDI4DataScience, NFDI4Ing, NFDI4Microbiota
ISE – Information Service Engineering, FIZ Karlsruhe, KIT, Karlsruhe	Knowledge Graphs and Linked Data Engineering, Knowledge Representations and Ontological Engineering, Hybrid AI	Prof. Dr. Harald Sack	NFDI4Culture NFDI4Chem MaRDI NFDI-MatWerk NFDI4DataScience NFDI4Memory NFDI4Objects FAIRagro NFDIxCS

Table 2: List of partner institutions joining as participants in the next funding phases

- Planned runtime of the project

1 year

- Summary of the proposal in English and German

English

Knowledge Graphs (KG) play an integral role in achieving interoperability in NFDI. While several consortia are building individual KG solutions, reusable KG infrastructure-as-a-service (KGI) is missing. To start discipline-specific KG work easily, NFDI consortia, participant institutions and researchers need reusable, scalable KGI. The proposed base service will provide it. Following a landscape analysis, the initialization phase will focus on a pilot KGI based on Wikibase,[1] the software behind Wikidata.[2] Wikidata is among the most popular, large-scale implementations of KG technology worldwide. It is widely used in scientific knowledge management and is an important advocacy tool for open data. Developing the pilot KGI includes establishing easy-to-scale infrastructure-as-a-service allowing NFDI stakeholders to create KGs without administrative overhead; developing an interoperability framework for connecting KGs with research infrastructures; establishing a KGI-consultancy to increase adoption of the KGI-service. The outcomes of the initialization phase ensure the expansion of the service in subsequent phases: the landscape analysis will lead to new partnerships with additional KG-tool providers; the consultancy service will expand operation beyond Wikibase KGs; and the interoperability framework will extend across Wikibase and non-Wikibase KGs towards a unified NFDI with a EOSC compatibility layer.

German

Wissensgraphen (Knowledge Graphs, KG) spielen eine wesentliche Rolle bei der Realisierung von Interoperabilität in der NFDI. Während mehrere Konsortien individuelle KG-Lösungen entwickeln, fehlt ein wiederverwendbarer KG-Infrastrukturdienst (KGI). Um mit der disziplinspezifischen KG-Arbeit problemlos beginnen zu können, benötigen NFDI-Konsortien, teilnehmende Institutionen und Forscher wiederverwendbare, skalierbare KGI. Der vorgeschlagene Basisdienst wird diese bereitstellen. Nach einer Landschaftsanalyse wird sich die Initialisierungsphase auf eine Pilot-KGI konzentrieren, die auf Wikibase basiert, der Software hinter Wikidata. Letzteres gehört zu den populärsten, groß angelegten Implementierungen der KG-Technologie weltweit. Wikidata ist im wissenschaftlichen Wissensmanagement weit verbreitet und stellt ein wichtiges Instrument zur Förderung offener Daten dar. Die Entwicklung der Pilot-KGI umfasst die Einrichtung eines einfach zu skalierenden Infrastrukturdienstes, der es den Mitgliedern der NFDI ermöglicht, KGs mit minimalem Aufwand zu erstellen. Ebenso schließt dieser die Entwicklung eines Interoperabilitätsrahmens für die Verbindung von KGs mit anderen Forschungsinfrastrukturen ein sowie die Einrichtung einer KGI-Beratungsstelle, um die Nutzung des KGI-Dienstes zu fördern. Die Ergebnisse der Initialisierungsphase ermöglichen die Erweiterung des Dienstes in den folgenden Phasen: Die Landschaftsanalyse wird zu neuen Partnerschaften mit zusätzlichen KG-Tool-Anbietern führen; der Beratungsdienst wird den Dienst über Wikibase-KGs hinaus ausweiten; und der Interoperabilitätsrahmen wird sich auf Wikibase- und Nicht-Wikibase-KGs erstrecken, um eine integrierte NFDI mit einer EOSC-Kompatibilitätsschicht zu schaffen.

[1] https://wikiba.se
[2] https://www.wikidata.org

2. State of the Art of Proposed Basic Service

Background and Motivation

Definitions

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[3] (RDF, SPARQL, etc.) and often used for agile data integration.

Knowledge Graph Infrastructure (KGI) in the context of this proposal is not only a triplestore or graph database. KGI is a whole ecosystem of software allowing to create a KG, including tools for data import, validation and export, collaborative frontends, search APIs and SPARQL endpoints with visualization, Extract-Transform-Load and data linking software adapted to the technology stack. An example is Wikibase KGI.

Motivation for KGI in NFDI

The Strategic Research and Innovation Agenda (SRIA) 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 Germany as a vehicle to facilitate data exchange between institutions and individual researchers. In addition to multilateral data- and knowledge-sharing within NFDI, linking domain-specific ontology-supported KGs with the multidisciplinary, crowdsourced knowledge in Wikidata KG will enable significant further applications. The linking between expert knowledge systems and world knowledge empowers lay persons to benefit from high-quality research data and ultimately contributes to increasing the faith in scientific research in society.

To date, several consortia have started building individual KG solutions or providing KG-compatible data using different formats and endpoints. In some cases, individual solutions are required to meet highly specific requirements. However, in many cases, the technical and organizational overhead to run these services can be hard to justify and is a burden to individual consortia or individual institutional members of consortia. KG infrastructure-as-a-service in NFDI is missing and needed.

The concept for a KGI service is motivated by ongoing developments across the NFDI Sections, which represent the most important cross-disciplinary common interests of NFDI consortia. A goal of Section "Training & Education" is the creation of a common platform for RDM teaching materials, while Section "Industry engagement" is aiming to create a joint exchange platform between industry and science. These sections can use the KGI service and address discipline-specific needs of their members using semantic web technologies. Furthermore, Sections "Common Infrastructures" and "Metadata, Terminologies, Provenance" are putting forward proposals for base services that can either contribute to enhancing the functionality of the KGI service or directly benefit from it. For instance, Identity and Access Management (IAM) services can ensure NFDI participants can seamlessly access the KGI. Persistent Identifiers (PID) and Terminology Services (TS) will contribute to the conformity of standard ontologies and identifiers reused by KGI across relevant disciplines. At the same time, KGI will ensure all data shared by NFDI participants has PIDs and follows the preferred ontologies specified for each domain via the TS. In addition, the Harvesting and Discovery Enhancing Service (HaDES)

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

proposal will benefit from the possibility of accessing data via standard APIs and SPARQL endpoints supported by the KGI service. Lastly, the Data Management Plans (DMP) service proposal aims to support research data compliance with FAIR principles and interoperability with national and international data services (NFDI Research Data Commons and EOSC), both of these goals are shared by the KGI proposal.

Motivation for Wikibase as KGI

Through the collaborative efforts of the sections, we have identified that several consortia and individual institutions already rely on a single KG service to meet a range of their domain-specific goals (Blümel, et al, 2022; Shigapov, 2022). That is Wikibase, the software behind Wikidata, the collaboratively curated knowledge graph in the open knowledge ecosystem around Wikipedia. Wikidata is already widely used across many domains of data, information and knowledge management, including scientometrics and academic research data management for the life sciences, computer science, as well as cultural heritage, among other fields (Rossenova, et al, 2022). The popularity of Wikidata – both as a repository to upload data to, and a rich resource on the linked open data (LOD) cloud to federate with – can be taken as 'proof of concept' that an approach to KGs that involves a mix of human- and machine-readable interfaces can lower the barrier to participation across a wide range of disciplinary fields and foster the creation of complex, cross-disciplinary connections. Furthermore, the identical approach to interfacing with data, modelling and querying data between Wikidata and Wikibase provides native federation capabilities across these services, which NFDI consortia that use Wikibase are already taking advantage of. In addition, a prototype Kubernetes orchestration for Wikibase deployment has already been released and is in beta testing by Wikimedia Germany.[4] Although this particular prototype does not meet all requirements for a research data management infrastructure, its open source development[5] will be a helpful model for developing the KGI service.

A dedication to develop Wikibase as an integral part of KGI for NFDI will streamline deployment and customization, and improve interoperability at scale. This work is intended to lessen the burden on individual consortia, use existing synergies, and provide ready-to-use infrastructure for KGs.

State of the art

The Working Group "Knowledge Graphs" (WG KGs) in Section "Metadata, Terminologies, Provenance" of the NFDI was set up to coordinate the development and use of KGs in all NFDI consortia. The initial member list of the group included 25 members representing 17 consortia. Consulting requests on building a KG in NFDI consortia regularly appear in the WG KGs mailing list (with 78 subscribers) and get discussed at the meetings of the WG.

Members of the WG have already started to evaluate the state of the art of KG adoption in NFDI and to discuss data integration and federation approaches. Several members of NFDI consortia presented different search technologies over distributed metadata. They involve both RDF knowledge graphs and PID graph technologies. The primary use cases of KGs in NFDI were also identified: 1) data integration across repositories, institutions, consortia and NFDI; and 2) providing access to (meta)data via SPARQL endpoints. Members of the WG have also prepared an overview survey for KG services in NFDI, which is planned to be implemented via WP1 of this proposal.

Some consortia have already carried out smaller-scale surveys that provide insights into a specific domain. For example, surveys organized by NFDI4Culture have gathered data from the

[4] https://www.wikibase.cloud
[5] https://github.com/wbstack

cultural heritage domain and showed low overall adoption of KGs.[6] This has prompted the creation of a culture-specific Linked Open Data Working Group,[7] which produced a comparison of LOD tools commonly used by culture partners (Nasarek, et al, 2023). Wikibase emerged as one of the key tools there since it not only provides a complete infrastructure stack, including SPARQL endpoints and well-documented APIs, but is supported by a mature open source community of users and developers – larger and more established compared to all other tools that were compared. Similar reasons led MaRDI to choose Wikibase as the underlying infrastructure for their KG data portal for the mathematical sciences. At BERD@NFDI the Wikibase KGs are used for integration of German (modern and historical) company data (Shigapov and Schumm, 2021). A deciding factor for choosing Wikibase there was also the well developed ecosystem of KG-supporting plugins, extensions and software.[8] Tools like OpenRefine and various named-entity-linking software adapted for Wikibase make agile data integration easier.[9]

Alongside its best known public instance – Wikidata – Wikibase as an infrastructure service can be comfortably evaluated as standing at TRL 9 (read more below), an actual system proven in (large-scale) operational environments.

Still, the survey planned within WP1 of this proposal will help to build a full picture over what currently is a fragmented landscape of individual consortia initiatives. Only after the landscape analysis is complete, we will be in a position to consider if and which additional tools and components may be required to become parts of the integration and ramping-up phases of the KGI service.

Own Preparatory Work for the Basic Service

Although a detailed set of requirements can only be drafted after the landscape analysis has been completed, some general **requirements** have already been identified following discussions within the Sections and representative members of the WG KG:

- Deployment scalability – multiple consortia and institutional partners need to be able to create and customize KGs with minimal administrative overhead and internal cost for specialist staff.
- Interoperability framework – easy-to-use and run data exchange pipelines are needed to facilitate interoperability as it is unrealistic to expect that all consortia and institutional partners will conform to a single (meta)data standard or ontology implementation.

These requirements point to **functional and operational gaps** within Wikibase, which can be addressed in the initialisation phase:

- Running a single Wikibase instance is computationally intensive and requires specialist knowledge: the current Wikibase.Cloud prototype aims to reduce computational resources required to run Wikibase instances in the hundreds, but problems with search indexing and data volume scaling across these instances are not yet solved.[10]
- There are no out-of-the-box interfaces for data import & export for Wikibase that comply with standards such as OAI-PMH. Data integration work is currently done piece-meal by individual projects and adjusted for every single use case, this is a hurdle for scalability of the service across multiple instances and knowledge domains.

---

[6] https://nfdi4culture.de/news-events/events/linked-open-data-meeting-knowledge-graph-strategy.html
[7] https://nfdi4culture.de/news-events/events/first-open-meeting-of-the-4culture-linked-open-data-working-group
[8] https://github.com/shigapov/wikibase-knowledge-graphs
[9] https://madoc.bib.uni-mannheim.de/59865
[10] https://meta.wikimedia.org/wiki/Wikibase/Wikibase.cloud#Updates

- A further challenge for achieving interoperability is the fact that Wikibase does not reuse existing RDF data easily and does not natively support standard ontologies. It is natively interoperable with Wikidata and other Wikibase KGs, but data mappings to other data sources are not easily generated or subsequently queried.

Open licenses and **existing standards**:

- The pilot KGI builds on existing, openly licensed technologies – Wikibase and Wikibase.Cloud, which are based on a number of established standards (RDF, SPARQL).
- Other data standards playing a role in the interoperability framework are openly licensed as well (e.g., OAI-PMH).
- Still, there is a necessity to develop better integration between other data sources and individual Wikibase instances, as well as facilitate easier integration and federation across several individual Wikibase instances and Wikidata.

Alternatives:

- Alternatives to the pilot KGI service can be observed across some consortia.[11] They develop individual solutions based on orchestration of multiple, non-natively integrated tools to perform the various tasks required for data modeling, data integration, storage, publication and querying. However, so far no services have been brought up at Section and WG KGs meetings that have reached high enough TRL with operationally proven ability to scale well across multiple knowledge domains.

Conflicting standards/implementations:

- N/A - the service is envisioned to be interoperable and possible to integrate easily with other related base services (e.g., IAM, TS, HaDES, PIDs, DMP).

Support from consortia:

The principal investigators of this proposal represent consortia that already work on Wikibase-related services.

- In the context of NFDI4Culture, TIB provides Wikibase Docker images which contain a number of modifications suited to the cultural communities.[12] In 2022, NFDI4Culture commissioned extensions for Wikibase which aimed to improve its compatibility with existing RDF ontologies,[13] and contributed to the maintenance and improvement of the OpenRefine Reconciliation service for individual Wikibase instances.[14] TIB is also the host institution for the ORKG[15] service, which is planned to be one of the additional services integrated following the initialisation phase.
- The KGs of German company data in BERD@NFDI are developed using Wikibase by UB Mannheim (Shigapov and Schumm, 2021). UB Mannheim has developed Wikibase-supporting software for data import (Shigapov, et al, 2021) and table

[11] Example from NFDI4Chem: https://github.com/StroemPhi/NMRspec
[12] https://gitlab.com/nfdi4culture/ta1-data-enrichment/wikibase-docker
[13] https://github.com/ProfessionalWiki/WikibaseRDF
[14] https://gitlab.com/nfdi4culture/ta1-data-enrichment/openrefine-wikibase and https://gitlab.com/nfdi4culture/ta1-data-enrichment/openrefine-wikibase-media
[15] https://orkg.org/about/1/Overview

annotation (entity linking, entity typing and relation extraction)[16] (Shigapov, et al, 2020). It has also commissioned a Wikibase export extension[17] for non-technical users.

- MaRDI is developing a KG for the mathematical sciences, for which Wikibase was already identified as the primary choice in the proposal. This Wikibase instance has been live since 2022. MaRDI contributed more than 100 code changes to the MediaWiki and Wikibase software and has thus improved the mathematics not only for the MaRDI portal but also to Wikipedia and Wikidata at the same time. Moreover, MaRDI decoupled the QuickStatements Tool from Wikidata so that smaller datasets in CSV files can not only be imported to Wikidata but also to other Wikibase instances such as the MaRDI portal. MaRDI has also deployed an instance[18] of the Wikibase data visualization frontend Scholia[19] and is adapting it to improve support for mathematical use cases. Further work is underway to adapt the Wikibase infrastructure to the needs of the mathematics community, e.g. by way of improving support for mathematical data types and formulas as well as import and export according to mathematical use cases.

Additional consortia (including Chem, DataScience, KonsortSWD, MicroBiota, Memory, Text+, Biodiversity) have also expressed interest in the service and willingness to participate in the follow up integration and ramping-up phases, helping to grow the adoption of the Wikibase KGI service and to identify additional tools and software packages that may be needed to become part of the core KGI service in order to meet diverse consortial requirements.

Current Technical Readiness Level /TRL) of the proposed Basic Service

For the purposes of NFDI, key aspects of the Wikibase ecosystem are situated between TRL7 and TRL9: Wikidata is a TRL9-level deployment that has been in operation for over ten years, handles 100 million entities, over ten thousand relationships, tens of thousands of monthly contributors and millions of daily queries. Several dozens of other Wikibase instances focused on specific knowledge domains have also been operational for several years at TRL9 levels, though at smaller scales. Most of them required some modifications and customizations – including the development of new extensions – that have largely reached TRL8 and TRL9 by now. Among them are Rhizome Artbase[20] for contemporary digital art, FactGrid[21] for historical research, Lingua Libre[22] for audiovisual data, OpenStreetMap Metadata[23] for metadata in OpenStreetMap[24], PersonalData.io[25] for a personal data ecosystem, EU knowledge graph[26] for entities about European Union-funded research, enslaved.org[27] of people of the historical slave trade, MiMoTextBase[28] for French Enlightenment novels.

Deployments in NFDI contexts require additional modifications and customizations, most of which are currently TRL7 and below. The community of Wikibase users and developers fosters regular interactions in multiple ways and a dedicated Stakeholder Group focusing on cross-institutional collaboration – including NFDI consortia representatives – has been operational since 2020.[29]

---

[16] https://github.com/UB-Mannheim/bbw
[17] https://github.com/ProfessionalWiki/WikibaseExport
[18] https://scholia.portal.mardi4nfdi.de/
[19] https://github.com/WDscholia/scholia
[20] https://artbase.rhizome.org
[21] https://database.factgrid.de
[22] https://linguali bre.org
[23] https://wiki.openstreetmap.org/wiki/Data_items
[24] https://www.openstreetmap.org
[25] https://wiki.personaldata.io
[26] https://linkedopendata.eu
[27] https://lod.enslaved.org
[28] https://data.mimotext.uni-trier.de
[29] https://wbstakeholder.group

3. SWOT Analysis

The following SWOT analysis is made for the pilot KGI service based on the Wikibase software.

Internal	**Strengths** 1. Based on TRL 9 software proven in operational environment with more than 100 million of entities (Wikidata). 2. Improving deployment scalability based on a technology proven to work (Kubernetes already used with Wikibase.Cloud, Wikidata and EU knowledge graph) will reduce adoption barriers such as high computing costs for individual installations and the need for specialist KGI staff at each institution. 3. Apart from the SPARQL endpoint, the Wikibase KGI provides a rich API and statement-level search via WikibaseCirrusSearch.[30] 4. Every entity is accessible via API in json, rdf, ttl, nt or jsonld formats. 5. Wikibase KGI includes out-of-the-box GUI for collaborative editing and discussions. 5. It stores the history of all changes. 6. It has many extensions and supporting tools for data validation, import, wrangling and reconciliation. 8. Wikibase KGs are adaptable to various domains, proven by operationally running instances.	**Weaknesses** 1. KG technology and SPARQL endpoints have a higher learning curve than RDM solutions based on relational databases. *Mitigation: More and more learning materials on KGs are available freely on the Internet: KIT Knowledge Graph course,[31] Stanford Knowledge Graphs course CS 520,[32] Knowledge Graphs E-book (Hogan, et al, 2022).[33]* 2. Standard ontologies cannot simply be imported into Wikibase. This results in the need for additional data modeling efforts and mapping to Wikibase's own RDF structure. *Mitigation: 1) Experimental work on importing standard ontologies into Wikibase was commissioned by NFDI4Culture and is available for testing as a Wikibase Extension.[34] 2) There is also a library of patterns that link between a traditional ontology design pattern and the underlying Wikibase data model (Eells, et al, 2021). 3) In the integration phase the KGI service can be supplemented with KG-software without such a weakness.* 3. Wikibase is by default an open system for data editing in a crowdsourced 'wiki'-style. This works well in the context of Wikidata which is not meant to be a primary research source, but does not necessarily fit the needs of researchers. *Mitigation: More granular identity and access management controls over editing rights will have to be implemented in later phases of the service development. These can benefit from the IAM base service being established first.*
External	**Opportunities** 1. Join an established open and international community of software developers and users working collaboratively to improve the ecosystem of Wikibase KGI (Pintscher et al, 2019). The PI of the lead institution has already organized the Wikibase Stakeholder Group which promotes sustainable development of the Wikibase software and extensions ecosystem. All PIs are members of the group. The third PI is the lead developer of mathematics support in Wikidata (Schubotz and Wicke, 2014) and	**Threats** 1. KG-software other than Wikibase (or its components) may emerge in the meantime and gain more popularity among users and service providers. *Mitigation: High-performance triplestores and graph databases which outperform Blazegraph are an opportunity rather than a threat, since they could be used as a replacement for Blazegraph.* 2. Other technical approaches for consortium-level data integration become favored over KGs.

[30] https://www.mediawiki.org/wiki/Extension:WikibaseCirrusSearch
[31] https://open.hpi.de/courses/knowledgegraphs2020
[32] https://web.stanford.edu/~vinayc/kg/
[33] https://kgbook.org
[34] https://github.com/ProfessionalWiki/WikibaseRDF

	organized the NFDI group[35] at Phabricator, which coordinates the technical Wikibase tickets relevant for NFDI. 2. Interoperate with the data in Wikidata easily, while avoiding the restrictions on primary research data enforced there. The second PI organized "WikiProject NFDI"[36] at Wikidata with the aim to create and edit entities and entity schemas relevant for NFDI. 3. Improve overall (meta)data quality based on automated services that can run over KGI. 4. Connect NFDI data to existing KG-based services that promote open research and machine-actionable data enrichment (e.g. ORKG). 5. Establish pipelines that streamline access to data from the local consortia level to the international level (e.g. EOSC, OpenAIRE).	*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.* 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) To reduce duplication in the efforts of consortia that already use Wikibase as a research data management service and to provide easy interlinking pathways to Wikidata.

(The motivations for defining this goal were already outlined in section 2 of this proposal.)

2) To support the provision of data via SPARQL endpoints across a wide range of RDM solutions already in place amongst consortia and their members.

(The motivation for defining this goal is a recognition of the diversity of use cases and solutions that are likely to surface in the landscape analysis. Therefore, this goal provides a flexible approach towards the service architecture of the KGI.)

To achieve these goals, the following **five objectives** have been identified:

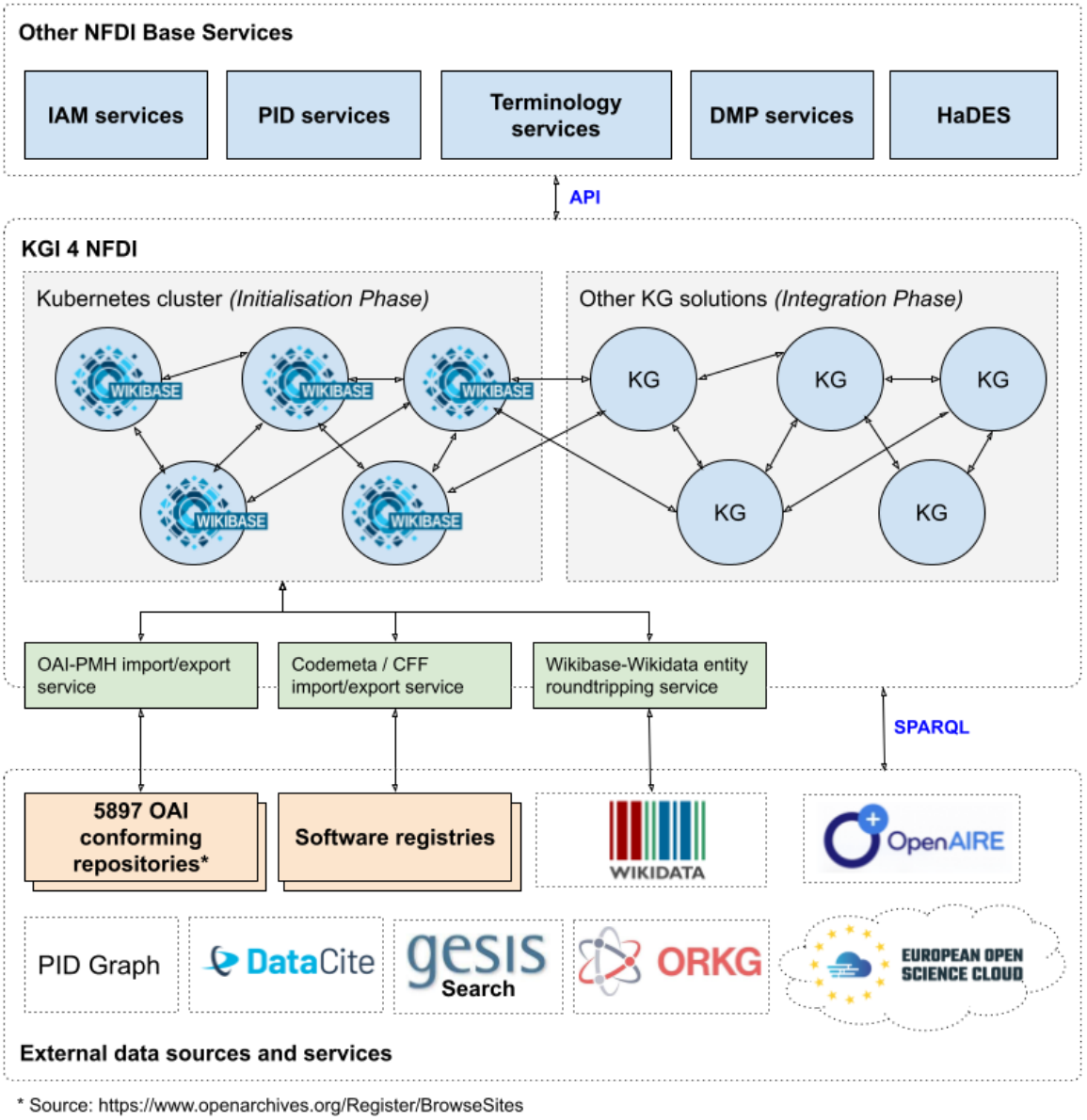
1. Optimize and scale up deployment of Wikibase instances with a service architecture that is easy to maintain and reuse within NFDI contexts.
2. Provide interoperability for Wikibase instances at different levels of the infrastructure:
    - between Wikibase instances and other scholarly infrastructures;
    - between NFDI Wikibase instances and Wikidata;
    - between different Wikibase instances deployed in NFDI contexts.
3. Collaborate with all relevant stakeholders in order to stay up to date with latest developments and requirements coming from the NFDI community, including commitment to:
    - carry out surveys;
    - participate in relevant Section meetings and Working Group meetings;
    - remain in consultation with stakeholder institutions working on KG services for NFDI;
    - stay informed of relevant national and international developments in the field of KGs beyond NFDI.

[35] https://phabricator.wikimedia.org/project/profile/5664
[36] https://www.wikidata.org/wiki/Wikidata:WikiProject_NFDI

4. Identify potential KGI service components (e.g., alternative triplestores, data integration APIs, etc.) that fall outside the boundaries of the Wikibase software ecosystem, but would nonetheless benefit NFDI consortia and their members; and incorporate these components into the service architecture.
5. Provide support to all consortia and their members in utilizing KGI services for their own use cases.

The **Initialisation phase** of the service will focus on achieving the first goal identified in the working concept and will do that through working packages designed to meet objectives 1-5.

In the **Integration and Ramping-Up Phases** we will continue to work on the first goal, but will also focus on achieving the second goal of the service. Working packages for these phases will focus on deliverables connected to objectives 3, 4 and 5.



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

Figure 1: Overview diagram of the working concept for the KGI service.

Service initialisation concept

Our service initialisation concept includes a landscape analysis, operationalizing a pilot KGI service based on the Wikibase software, and establishing a consultancy to support onboarding as well as service alignment processes across the NFDI communities.

This will be carried out through four working packages (WPs):

1. Landscape analysis and requirements gathering: This WP will include identifying the needs of various NFDI communities with respect to knowledge graphs through a survey developed with the WG KGs in NFDI Section Metadata. The survey will target KG experts working in the relevant task areas within each consortia – we expect they will support this effort since the intended analysis falls within the remit of regular KG task area activities. The documentation and analysis of survey results will directly inform how we devise an agile plan to continuously adapt the KGI service concept to better fit the needs of diverse NFDI communities. A part of this plan will also include identifying alternative software components that fall outside the boundaries of the Wikibase software ecosystem and reporting on their applicability to NFDI contexts. The survey and the participation of this proposal's PIs into the WG KG will aim to assist continuous communication channels with KG experts in most or all NFDI consortia. The concept of this WP has efficient mechanisms for gathering expert information from consortia that can directly feedback into the development of the KGI service throughout all its phases.

2. Deployment service: This WP will be responsible for developing and setting up the service architecture for the knowledge graph infrastructure based on Wikibase software (TRL=9). The work will focus on setting up a Kubernetes cluster orchestration at TIB in close collaboration with the supporting partner – Wikimedia Germany (WMDE). The WP will build upon the prototype work already underway at WMDE for their Wikibase.Cloud product. What this WP will secure is stability in service provision and adapting the public, non-specialist Wikibase.Cloud configuration to a service that meets specialist research needs and includes all the previously commissioned extensions and customizations carried out by the participating institutions (cf. Section 2).

3. Interoperability service: This WP will aim to significantly ease existing workflows when it comes to data integration between various RDM systems and Wikibase instances (including Wikidata). An initial focus will be the provision of services for established data exchange formats that are already commonly used by consortia (as evidenced through discussions in relevant working groups of Section Metadata such as the 'Search and Harvesting' and 'Ontology mapping and harmonization' WGs), e.g., OAI-PMH for bibliographic data and CodeMeta / CFF for software metadata. Crucially, a task within WP will be securing easier interoperability between individual Wikibase instances and Wikidata – a goal for all consortia members already using Wikibase and/or Wikidata as data portals. The outcomes of this WP are essential for securing interoperability between NFDI data (including other base services, e.g., HaDES) and data in open public knowledge platforms including international initiatives (e.g., EOSC).

4. Consultancy: This WP will establish direct communication between the base service and representative partners in all NFDI consortia, relevant Sections and Working Groups. We will be providing consultancy and support for KG building with the KGI service; while also gathering feedback on how onboarding can be improved and how service alignment processes can be better facilitated in order to grow KG adoption across all consortia.

At the end of the initialisation phase, the specific requirements by all relevant consortia, institutions and researchers for KGI services will be clarified and synthesized into the detailed proposals for the next phases. Furthermore, we hope to use the pilot KGI as a test case in

terms of how other software components (beyond Wikibase) can also be effectively deployed and made interoperable within the overall KGI service architecture. By working with a software with a high TRL, we hope to move through deployment and integration work packages without significant hurdles and with plenty of time for extracting best practices that can be applied in later phases of the service development.

Integration and Ramping-Up-For-Operation outlook

Integration phase concept: The KGI service shall be adapted to the needs of NFDI clarified in the initialisation phase. We expect that the pilot KGI service based on Wikibase software might not meet all the needs of NFDI consortia and have already contacted potential partners for the integration phase (see table of partner institutions – GESIS, ZBMED, Fraunhofer FOKUS, ORKG, ISE FIZ Karlsruhe). Partners can contribute alternative, open-source, reusable KGI solutions which may or may not be based on the Wikibase software ecosystem. Alternative solutions will be expected to have high TRL (7 or above) and to foster additional capacity building for scalability and interoperability based on lessons learned from the initialisation phase. The pilot KGI service based on Wikibase software will also continue to be developed further in accordance with the needs of NFDI. Additional work packages that can be considered include: integrating identity and access management for the KGI service based on the corresponding base service (IAM); integrating quality assurance mechanisms and components for ontology reusing, repairing and enrichment (e.g., Terminology Service). Requirements gathering activities and consulting services will be continuously carried out throughout this phase as well.

Ramping-Up-For-Operation phase concept: The base KGI service based on a range of KG-software solutions shall be reusable and fully operational. Thanks to the landscape analysis and requirements gathering in the initialisation phase, the continuous agile development and the updates to the service made in the integration phase, the KGI service will strive to meet requirements of all NFDI consortia, institutions and researchers. Ongoing consulting services and facilitating connection to international initiatives like EOSC will also be prioritized during this phase.

Risks and challenges

Risks and challenges are described in the table with SWOT analysis. We outline here the general risks and challenges, addressing the three phases of the project (initialisation, integration and ramping-up-for-operation) and the criteria[37] of selection provided by Base4NFDI.

1. Getting support of NFDI consortia: approval of 25%, 50% and 75% (“potentially all”) consortia in the corresponding phases. It might be a challenge to get support from "potentially all" consortia. A risk of not getting such support exists.
Risk-mitigation measures: The needs of all consortia shall be clarified during landscape analysis in the initialisation phase. All needs of NFDI consortia must be met in the integration and ramping-up-for-operation phases.

2. Interoperability in NFDI: the KGI service must be interoperable with NFDI services. The challenge is that we do not have yet an overview of all potential NFDI services. There is a risk that interoperability with a certain NFDI service was not considered in the initialisation phase.
Risk-mitigation measures: WP1 "Landscape analysis", WP3 "Interoperability", WP4 "Consultancy", feedback from consulting institutions, working groups and sections.

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

3. International interoperability: ability to integrate with other national and international infrastructures, in particular EOSC. Though the KGI service aims by design for international interoperability, there is a risk that certain national and international infrastructures are not in our focus in the initialisation phase.
Risk-mitigation measures: WP1 "Landscape analysis", WP3 "Interoperability", WP4 "Consultancy", involvement in EOSC, feedback from consulting institutions, working groups and sections.

4. Recruiting qualified staff: ability to recruit highly specialized staff within the timeframe of the WPs and schedule of Deliverables. Knowledge Graph technologies, and specifically the pilot case software Wikibase, require niche expertise and there are already staff shortages noted within one of the partner organizations, Wikimedia Germany.
Risk-mitigation measures: The PIs are members and co-facilitators of the Wikibase Stakeholder Group which boasts a large international community of Wikibase experts and has been a successful community for recruitment in the past (e.g. within TIB's team). Additionally, the consortia and member institutions leading this proposal can rely on their own high visibility and reputation within the German research ecosystem for recruiting scientific staff.

5. Work Programme

5.1. Overview of work packages

Work package	Deliverables (D) and milestones (M)	Responsible partner(s)
1. Landscape Analysis	D1.1 Document with results of the survey D1.2 Report with overviews of knowledge graphs for RDM, KGIs-as-a-service & KGs in NFDI D1.3 Document with requirements for the service M1.1 Requirements for the KGI service in NFDI are clarified and documented	TIB, FIZ, UBM
2. Deployment scalability	D2.1 High-availability, internally-maintained Kubernetes cluster D2.2 Resource configuration to match requirements of Wikibase software D2.3 Maintenance pipelines to manage Wikibase nodes at scale D2.4 Stress test system to assess concrete scalability expectations M2.1 The pilot KGI service is deployed and running	TIB
3. Interoperability	D3.1 Converter OAI-PMH to Wikibase to OAI-PMH D3.2 Converter CFF-Wikibase-CFF D3.3 Converter CodeMeta-Wikibase-CodeMeta D3.4 Identifier converter NFDI Wikibase to and from Wikidata M3.1 Pilot KGI interoperability pipeline	FIZ
4. Consultancy	D4.1 KGI consultancy with consultations per emails and video calls D4.2 Document with feedback and needs of users and providers of the KGI service D4.3 Documentation for the KGI service D4.4 Readiness to consult about KGIs beyond Wikibase M4.1 The KGI consultancy is running M4.2 The KGI consultancy is extended to KGI services beyond Wikibase	UBM

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

5.2. Detailed work programme

Working Packages, Tasks, Deliverables and Milestones	Base Service “Knowledge Graph Infrastructure”											
	1	2	3	4	5	6	7	8	9	10	11	12
WP1: Landscape Analysis									M1.1			
Task 1.1 (D1.1)				D1.1								
Task 1.2 (D1.2)								D1.2				
Task 1.3 (D1.3)									D1.3			
WP2: Deployment scalability												M2.1
Task 2.1 (D2.1, D2.2)				D2.1		D2.2						
Task 2.2 (D2.3)									D2.3			
Task 2.3 (D2.4)									D2.4			
WP3: Interoperability												M3.1
Task 3.1 (D3.1)								D3.1				
Task 3.2 (D3.2, D3.3)				D3.2		D3.3						
Task 3.3 (D3.4)												D3.4
WP4: Consultancy			M4.1									M4.2
Task 4.1 (D4.1, D4.2)			D4.1						D4.2			
Task 4.2 (D4.3)									D4.3			
Task 4.3 (D4.4)												D4.4

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

5.2.1 WP1: Landscape Analysis (Month 1-9)

Who is using KGs in NFDI? What kind of infrastructures, software and ontologies are they based on? What are the use cases in different consortia? What are the blocking factors for wider adoption of KG services?

Task 1.1 Survey on KGs in NFDI

The NFDI Working Group "Knowledge graphs" has developed a survey on KGs in NFDI consortia using a Lime instance at UB Mannheim. The survey aims to identify all KG instances in NFDI (APIs, SPARQL endpoints, software and ontologies), gather best practices and blocking factors for implementing KGs and clarify the use cases of KGs. The needs of NFDI consortia, sections, institutions and individual researchers in the German research landscape with respect to KGs shall be also clarified.

Deliverable 1.1 Document with results of the survey (Month 4)

Task 1.2 Overviews of KGs for RDM, KGIs-as-a-service & KG portals in NFDI
The results of the survey need to be processed, analyzed and compared with the European and international KG-based research data infrastructures. Overviews of KGs for research data management tasks and infrastructure-as-a-service solutions shall be compiled.

Deliverable 1.2 Report with overviews of KGs for RDM, KGIs-as-a-service & KG portals in NFDI (Month 8)

Task 1.3 Requirements gathering for adjusting the service in the integration phase
If the needs of NFDI consortia, sections, institutions and individual researchers with respect to KGs are not fully met by the proposed base service, changes to the service will be proposed and documented in the draft of the proposal for the integration phase.

Deliverable 1.3 Document with requirements for the service (Month 9)

WP1 Milestones
M1.1 Requirements for the KGI service in NFDI are clarified and documented (Month 9)

5.2.2 WP2: Deployment scalability (Months 1-12)

What are the challenges involved in hosting a Kubernetes cluster for a multi-node Wikibase deployment? What systems can be used to enable maintenance at scale and avoid individual server configuration? What limits can be reasonably expected in supporting large datasets?

Task 2.1 Deploy Kubernetes cluster
Having identified Kubernetes as an appropriate and widely-supported platform on which to deploy a multi-node Wikibase ecosystem, work will be undertaken to investigate and implement a local deployment of such a system. This includes the configuration of resources (e.g. time-out limits, user rights, production-versus-test environments) to reflect the operational, computational and policy needs of Wikibase, notably prioritizing responsive querying, responsible contributions and uninterrupted services.

Deliverable 2.1 High-availability, internally-maintained Kubernetes cluster (Month 4)
Deliverable 2.2 Resource configuration to match requirements of Wikibase software (Month 6)

Task 2.2 Ansible, CD/CI DevOps to support mass node (or "pod") updates and maintenance
As Wikibase is not a singular entity, but a collection of open source tools, services and extensions, it requires a complicated update pattern. To avoid extensive work on individual server maintenance, work should be undertaken to develop pipelines for processing nodes in batch operations. This also entails respecting a level of environment conformity between instances (see also WP3), to prevent issues with nodes developing unusual characteristics, which could lead to unexpected behaviors during routine operations.

Deliverable 2.3 Maintenance pipelines to manage Wikibase nodes at scale (Month 9)

Task 2.3 Stress testing nodes to determine expected scalability
Regarding raw graph size, Wikidata is a Wikibase which significantly exceeds all other deployed community instances. A crucial stage in offering a reusable and extendable environment is testing and identifying clear limits of the quantity of data statements, change rates, number of contributors or volume of queries (among other parameters) which can be incorporated in nodes

before compromising performance. This should comprise both assessing Wikibases which rely only on data statements, and also instances which are "media-heavy" (i.e. incorporating media files), as these carry their own resource considerations.

Deliverable 2.4 Stress test system to assess concrete scalability expectations (Month 9)

WP2 Milestones

M2.1 The pilot KGI service is deployed and running (Month 12)

5.2.3 WP3: Interoperability with standard RDM-interfaces (e.g., OAI-PMH), Wikidata and beyond (Months 1-12)

How can data and metadata be integrated between NFDI KGs? How can data integration be enabled between NFDI KGs and other research infrastructures? Which mechanisms can be leveraged for data integration between Wikidata and NFDI KGs?

Task 3.1 Converting OAI-PMH to and from RDF/ SPARQL

Multiple NFDI consortia are operating OAI-PMH endpoints for metadata, as do many other infrastructures. These endpoints can not be queried via graph query languages like SPARQL, so in order to make the underlying data accessible in KGs, conversion workflows need to be developed. These should be bidirectional, so that bibliographic metadata already residing in KGs can also be integrated into OAI-PMH endpoints and thus exposed to harvesters.

Deliverable 3.1 Service for exporting/ importing OAI-PMH compliant data out of and into Wikibase is deployed (Month 8)

Task 3.2 Converting CodeMeta/ Citation File Format to and from RDF/ SPARQL

Software is increasingly recognized as an integral part of research data workflows. Metadata about software can be expressed using the emerging CodeMeta and CFF (Citation File Format) standards. Analogous to the OAI-PMH case above, integration of software metadata into KG-based workflows would benefit from robust conversion workflows to and from RDF/ SPARQL.

Deliverable 3.2 Service for importing and exporting CFF compliant data into and out of Wikibase via RDF and SPARQL is deployed (Month 4)
Deliverable 3.3 Service for importing and exporting CodeMeta compliant data into and out of Wikibase via RDF and SPARQL is deployed (Month 6)

Task 3.3 Roundtripping between Wikidata and Wikibase via external identifiers

Identifier mapping between entities in Wikidata and other Wikibase instances can be performed directly, but since these identifiers usually differ between Wikibase instances, an indirect approach using well-established external identifiers is often more efficient and practical. Based on the survey results (cf. D1.1), a set of external identifiers commonly used across NFDI KGs will be identified, for which a service using them to convert NFDI Wikibase identifiers to and from Wikidata identifiers will be developed.

Deliverable 3.4 Service to convert NFDI Wikibase identifiers to and from Wikidata identifiers via external identifiers is deployed (Month 12)

WP3 Milestones

M3.1 The pilot KGI interoperability pipelines are developed and documented (Month 12)

5.2.4 WP4: Consultancy (Months 1-12)

To increase adoption of the KGI service among users and service providers in NFDI, the KGI consultancy will be launched. The KGI consultant will onboard the users and service providers to the KGI, improve documentation for the KGI service, and document further feedback and service requirements.

Task 4.1 Creating the KGI consultancy

The KGI consultancy is aimed to help users and providers of the KGI services. The KGI consultant shall be available via emails and video calls. Users and providers of KGI services need different degrees of onboarding – from advice on general KG creation and use-case specification, to detailed instruction concerning specific infrastructural requirements and configurations. The KGI consultant shall develop adequate onboarding processes to suit multiple scenarios. The KGI consultant will document feedback and additional requirements of NFDI consortia, institutions and researchers with respect to the KGI service and provide them as supplementing documentation to WP 1 (Deliverable 1.3).

Deliverable 4.1 KGI consultancy with consultations per emails and video calls (Month 3)
Deliverable 4.2 Document with feedback and needs of users and providers of the KGI service (Month 9)

Task 4.2 Documenting the KGI service

In WPs 1, 2 and 3 large amounts of documentation will be created. The documentation should be collected, systematized, enhanced and presented in a user-friendly, publicly-available way. Both users and providers of the KGI service shall be able to find relevant information about the service.

Deliverable 4.3 Documentation for the KGI service (Month 9)

Task 4.3 Expanding consultancy to KGI services beyond Wikibase

The pilot KGI service might not meet the requirements of all consortia, institutions and researchers. These requirements will be documented in the Deliverable D1.3 (Month 9). One of the expected requirements is the ability 1) to reuse an existing ontology and to create a KG based on it, and 2) to embed existing RDF-data into the KGI service. The KGI-consultant shall have a good overview of the alternative open-source KG software and be ready to provide consulting beyond the pilot KGI service based on Wikibase.

Deliverable 4.4 Readiness to consult about KGIs beyond Wikibase (Month 12)

WP4 Milestones

M4.1 The KGI consultancy is running (Month 3)
M4.2 The KGI consultancy is extended to KGI services beyond Wikibase (Month 12)

III Appendix

a) Bibliography and list of references

Ina Blümel, Paul Duchesne, **Lozana Rossenova**, and Harald Sack. 2022. NFDI InfraTalk: Wikibase - knowledge graphs for RDM in NFDI4Culture. (7 March 2022). URL: https://www.youtube.com/watch?v=RPMkuDxHJtI&ab_channel=NFDIDirektorat.

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

Andrew Eells, Cogan Shimizu, Lu Zhou, Pascal Hitzler, Seila Gonzalez, and Dean Rehberger. 2021. Aligning Patterns to the Wikibase Model. Workshop on Ontology Design and Patterns 2021. URL: https://ceur-ws.org/Vol-3011/paper2.pdf

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

David Fichtmueller. 2021. Using Wikibase as a Platform to Develop a Semantic Biodiversity Standard. In: *1st NFDI Wikibase Workshop, online*. URL: https://docs.google.com/presentation/d/1i91OB9xPZVVovd8c7Cm2sOglQLM8CEeZed8grdVaFwU/edit.

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

Jose Emilio Labra Gayo, et al. 2021. Representing the Luxembourg Shared Authority File based on CIDOC-CRM in Wikibase. In: *SWIB 2021, online, 2021.* URL: https://swib.org/swib21/slides/05-03-gayo.pdf.

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

Lydia Pintscher, et al. 2019. Strategy for the Wikibase Ecosystem. URL: https://upload.wikimedia.org/wikipedia/commons/c/cc/Strategy_for_Wikibase_Ecosystem.pdf.

Andrea Piscopo. 2019. Structuring the World's Knowledge: Socio-Technical Processes and Data Quality in Wikidata. PhD thesis, University of Southampton, UK.

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

Philipp Scharpf, **Moritz Schubotz,** and Bela Gipp. 2021. Mathematics in Wikidata. *2nd Wikidata Workshop (Wikidata 2021) co-located with the 20th International Semantic Web Conference (ISWC 2021).* URL: https://doi.org/10.34657/8001

Philipp Scharpf, **Moritz Schubotz**, and Bela Gipp. 2018. Representing Mathematical Formulae in Content MathML using Wikidata. *3rd Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL 2018) co-located with the 41st International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2018).* URL: https://ceur-ws.org/Vol-2132/paper5.pdf

Moritz Schubotz, Abdou Youssef, Volker Markl, and Howard Cohl. 2015. Challenges of Mathematical Information Retrieval in the NTCIR-11 Math Wikipedia Task. *SIGIR '15: Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval.* URL: https://doi.org/10.1145/2766462.2767787x

Moritz Schubotz and Gabriel Wicke. 2014. Mathoid: Robust, Scalable, Fast and Accessible Math Rendering for Wikipedia. *Intelligent Computer Mathematics. CICM 2014.* URL: https://doi.org/10.1007/978-3-319-08434-3_17

Renat Shigapov. 2021. RaiseWikibase: Towards fast data import into Wikibase. In: *2nd Workshop on Wikibase in Knowledge Graph based Research Data Management (NFDI) Projects.* URL: https://madoc.bib.uni-mannheim.de/60059.

Renat Shigapov, et al. 2020. bbw: Matching csv to Wikidata via meta-lookup. *SemTab 2020: Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching co-located with the 19th International Semantic Web Conference (ISWC 2020).* Vol. 2775. RWTH Aachen. URL: http://ceur-ws.org/Vol-2775/paper2.pdf.

Renat Shigapov, Jörg Mechnich, and Irene Schumm. 2021. RaiseWikibase: Fast inserts into the BERD instance. *The Semantic Web: ESWC 2021 Satellite Events: Virtual Event, June 6–10, 2021, Revised Selected Papers 18.* Springer International Publishing. URL: https://doi.org/10.1007/978-3-030-80418-3.

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

Philip Strömert, and Oliver Koepler. 2022. Ontologies4Chem: A use case to build a NMR research data knowledge graph. *ACS Spring Meeting. 23.03.2022*. URL: https://docs.google.com/presentation/d/1qr2OiFVW4u-KFjtD71D08zOeEdH1q-nh/edit#slide=id.p1

Katherine Thornton, Kenneth Seals-Nutt, Euan Cochrane, and Carl Wilson. 2018. Wikidata for Digital Preservation. In: *Proceedings of iPRES'18, Cambridge, MA, USA, September 24–27, 2018.*

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