Proposal

Terminology Services 4 NFDI

TS4NFDI

A cross domain service for the provision, curation, development, harmonisation, and mapping of terminologies

Initialisation Phase

1. General Information

Lead institution

TIB – Leibniz Information Centre for Science and Technology (lead)

Welfengarten 1 B, 30167 Hannover

ZB MED – Information Centre for Life Sciences (co-lead)

Gleueler Straße 60, 50931 Cologne

Name of lead institution principle investigator

Dr. Oliver Koepler (oliver.koepler@tib.eu) (lead)

Roman Baum (baum@zbmed.de) (co-lead)

Participating institutions

Principal Investigator	Institution, location	Contact Email	Member in [consortium][1]
Dr. Oliver Koepler	TIB - Leibniz Information Centre for Science and Technology	oliver.koepler@tib.eu	NFDI4Chem
Roman Baum	ZB MED - Information Centre for Life Sciences	baum@zbmed.de	NFDI4Health, FAIRagro
Dr. Naouel Karam, Prof. Adrian Paschke	Institut für Angewandte Informatik (InfAI) e.V.	karam@infai.org paschke@infai.org	NFDI4Biodiversity

Table 1: List of participating institutions

Partner institutions without funding

Contact person	Institution, location	Contact E-mail	Member in [consortium]
Jakob Voss	Verbundzentrale des GBV (VZG), Göttingen	jakob.voss@gbv.de	NFDI4Objects, NFDI4Memory
Björn Usadel	Forschungszentrum Juelich GmbH, Juelich	b.usadel@fz-juelich.de	DataPLANT
Felix Engel	TIB - Leibniz Information Centre for Science and Technology, Hannover	Felix.Engel@tib.eu	NFDI4Ing

Table 2: List of partner institutions without funding

[1] Name one DFG consortium the institution is or has a route to become a member of and through which funds should be appropriated if this proposal is approved.

Planned duration of the project: 1 year

Summary of the proposal in English

Terminologies play a critical role in semantically linking research data management (RDM) across disciplines. They provide the basis for consensus definitions of entities, thereby ensuring conceptual alignment, even when the nomenclature differs between domains. The use of standardised terminologies promotes interoperability and data integration, as data described through common terminologies can be understood and used across different sources and disciplines significantly improving their reusability. Currently, NFDI consortia are working with terminologies at different stages of maturity and FAIRness. Terminologies are used in various services and tools to annotate and link research. To support these tasks, several terminology services (TSs) have been developed to publish, provide, describe, and curate terminologies. Their alignment, management, curation, and further development are challenging but necessary tasks. They involve consensus building and harmonisation across disciplines. The envisaged Terminology Services 4 NFDI (TS4NFDI) aims to standardise and harmonise terminology management within the NFDI, thereby facilitating consensus-building and interoperability of services across disciplines to achieve a shared knowledge representation and knowledge engineering framework. The service seeks to integrate and converge individual solutions into a standardised, interoperable, and sustainable service suite with Service Wrapper, API Gateway, mapping service, and re-usable GUI widgets.

Summary of the proposal in German

Terminologien spielen eine wichtige Rolle bei der semantischen Vernetzung von Forschungsdaten über Disziplinen hinweg. Sie bilden die Grundlage für die Entwicklung eines gemeinsamen Verständnisses von der Bedeutung von Entitäten, selbst wenn die Benennung von Entitäten in den Wissensdomänen abweicht. Die Verwendung standardisierter Terminologien verbessert die Interoperabilität und Datenintegration. Mit Terminologien beschriebene Daten können aus unterschiedlichen Quellen und Disziplinen übergreifend verstanden und genutzt werden, was ihre Wiederverwendbarkeit erheblich verbessert. Derzeit arbeiten NFDI-Konsortien mit Terminologien unterschiedlicher Reife und FAIRness. Diese Terminologien werden in verschiedenen Diensten und Werkzeugen verwendet, um Forschungsdaten zu annotieren und zu verknüpfen. Zu diesem Zweck wurden Terminologiedienste (TS) entwickelt, um Terminologien zu veröffentlichen, bereitzustellen, zu beschreiben und zu kuratieren. Ihr Management, Kuratierung und Weiterentwicklung sowie gemeinsame Ausrichtung sind herausfordernde, aber notwendige Aufgaben. Sie umfassen insbesondere auch die inhaltliche Konsensbildung und Harmonisierung über Disziplinen

hinweg. Der angestrebte Terminologie Basisdienst (TS4NFDI) hat das Ziel, das Terminologiemanagement innerhalb der NFDI zu standardisieren und zu harmonisieren und damit die Konsensbildung und Interoperabilität der Dienste über Disziplinen hinweg zu fördern, um einen gemeinsamen Rahmen für Wissensrepräsentationen aufzubauen. Der Service zielt darauf ab, individuelle Lösungen in eine standardisierte, interoperable und nachhaltige Service-Suite mit Service Wrapper, API Gateway, Mapping Service und wiederverwendbaren GUI-Widgets zu integrieren.

2. State-of-the-Art of Proposed Basic Service

2.1 Background

A Terminology Service (TS) is a web-based service to manage and provide access to terminologies (vocabularies, terminologies, ontologies, and classifications or thesauri). Functionalities of a terminology service include queries, browsing, visualisations and mappings between terminologies. Terminology Services (TSs) are used by various stakeholders with different expertise levels in different use cases. Ontology experts, knowledge workers and researchers may use a TS to manage, curate or design terminologies or search for terminologies or terms to annotate digital objects. Besides human users a TS can be used from other services like electronic lab notebooks, data repositories or annotation tools via APIs. These services embed terminologies in their user interfaces and workflows. Due to the crucial role of terminologies in annotating and linking research data across domains, several NFDI consortia are already utilising TSs and integrating them into their tool landscape for designing, curating, extending and linking terminologies and annotating (meta)data.

2.2 State-of-the-art

A number of technically mature TSs and repositories are publicly available. Terminology registries such as the Basic Register of Thesauri, Ontologies & Classifications (BARTOC) represent the simplest form of such a service [1]. Registries typically list comprehensive metadata about vocabularies, terminologies and ontologies and link to their original source while terminology repositories contain full terminology data in machine-readable form and, in addition, provide access to their data via APIs [2]. Well used TS frameworks like OntoPortal [3] and Ontology Lookup Service (OLS) [4] offer extended features for searching and browsing within the hosted terminologies and provide visualisation of hierarchical relationships in so-called tree views as well as information about concepts and properties. OntoPortal is a portal architecture developed by the OntoPortal Alliance [5] that can be customised for a domain to provide discipline-specific advanced semantic repositories. It has evolved from BioPortal, which

provides access to a large collection of biomedical ontologies and terminologies [6]. The Ontology Lookup Service, developed and hosted at the EMBL European Bioinformatics Institute (EBI), provides ontology search and visualisation services as well as data access and search via an API [4]. EBI also hosts the Ontology Xref Service (OxO) for finding mappings between terms from ontologies, importing the mappings from OLS and other terminology services [7]. There is also Skosmos, an open source web-based browser and publishing tool [8] specialised for Simple Knowledge Organisation System (SKOS) vocabularies [9]. It provides a user interface for browsing and searching the data as well as Linked Data access with APIs that support term-based searches. Further TSs include DANTE via JSKOS API [10] or the Linked Open Vocabularies (LOV) via SPARQL API [11], the General Multilingual Environmental Thesaurus (GEMET) [12], the GND (Gemeinsame Normdatei) [13] and the geographical database GeoNames [14].

2.3 Status of work results in preparation for the Basic Service

The integration of terminologies into the service landscape of the NFDI consortia involves two perspectives: a substantive engagement with terminologies including interdisciplinary consensus-building for the joint alignment of terminologies or their mappings. Secondly, the application of services and tools in finding, browsing and managing terminologies. These perspectives are addressed by two working groups: Ontology Harmonisation and Mapping (Onto-WG) and Terminology Services (WG-TS). The former records terminologies used in the consortia and strategizes harmonisation and mapping, while the latter identifies tools for managing terminologies. Both have compiled overviews of a) existing terminologies and b) terminology services used in the NFDI. Several Several consortia are using or intend to use the open source software of OLS (e.g. NFDI4Chem, NFDI4Health, and NFDI4Ing), OntoPortal (e.g. NFDI4Biodiversity, NFDI-MatWerk) and Skosmos (e.g. KonsortSWD, NFDI4Earth) for their domain-specific services. In addition, there are a number of services that have been developed to support the annotation of research data using terminologies specifically for their domain (e.g. DataPLANT, FAIRmat, NFDI4Objects, and Text+). In February 2023 65% of the consortia expressed their interest in a basic TS. A first version of a Metadata Annotation Workbench for NFDI4Health has been developed by ZB MED, integrating terminology services. In NFDI4Chem electronic lab notebooks and data repositories will retrieve terminologies for data annotation via a TS API. In addition, DataPLANT combines crowdsourced terminology extension and curation in its software solution SWATE [15], and mediates the changes made by the community to the nearest existing ontology. The current overview does not fully cover possible services and requirements of the third-round consortia yet. These numerous activities illustrate on one hand the need for terminologies and TSs, while on the other hand, they also reveal the risk of parallel,

incompatible individual service developments and redundant efforts across many consortia. Undoubtedly domain-specific needs have to be addressed. Nevertheless, a basic TS will contribute to harmonisation efforts, interdisciplinary reuse of terminologies and efficient data integration. It will provide standardised APIs for all NFDI services to access all available terminologies and enable the reuse and simple integration of core UI widgets to be implemented in consortia web interfaces.

2.4 Motivation

Numerous NFDI consortia are developing community solutions with custom-made terminology collections. While these solutions are crucial for addressing specific needs within respective disciplines, they often lead to a somewhat fragmented and siloed approach to terminology management and provision.

Moreover, while the content of terminology collections differs according to the scientific domains, the technical solutions and user interfaces often bear similarities. This represents a missed opportunity for synergy and improved efficiency. Furthermore, discipline-specific collections, while serving the needs of their respective fields, can inadvertently restrict the view on other, neighbouring disciplines.

The proposed basic Terminology Service TS4NFDI aims to standardise and harmonise collaborative terminology management within the NFDI. It provides a way to integrate and converge individual solutions into a standardised, interoperable, sustainable service suite embedded in a long-term NFDI infrastructure, thereby improving the overall efficiency and effectiveness of terminology management and use, as well as cross-domain collaboration within the NFDI community.

To achieve this goal TS4NFDI will provide the following features:

- Interconnection of existing Terminology Services (TSs) via **Service Wrapper** and the corresponding **API Gateway**. This allows for terminology use and management across disciplines, serving as a basis for terminology harmonisation efforts (Integration and Interoperability).
- **API Gateway**: Standardised, harmonised and interoperable access to Terminologies across all integrated Terminology Services, available for all NFDI-services.
- Central **Mapping Service** to create, manage and provide cross-domain mappings for terminologies (via SSSOM [16]) as a central service.
- **GUI widgets**: TS4NFDI offers widgets to be embedded into user interfaces to simplify the development of UIs and the visualisation of semantic information in other NFDI

services. This streamlines the process of integrating standard vocabularies and performing semantic annotation of data.

TS4NFDI fosters consensus finding and harmonisation across disciplines by a mapping service, UI widgets, and workflow tools to curate, align and harmonise ontologies and their application. As harmonisation progresses, domain ontologies may be aligned to common upper-level ontologies that can be provisioned centrally. The TS4NFDI architecture is able to integrate terminology collections into a central backend infrastructure ensuring the long-term availability of individual solutions. This strategy aims to maintain and provide continuity of service.

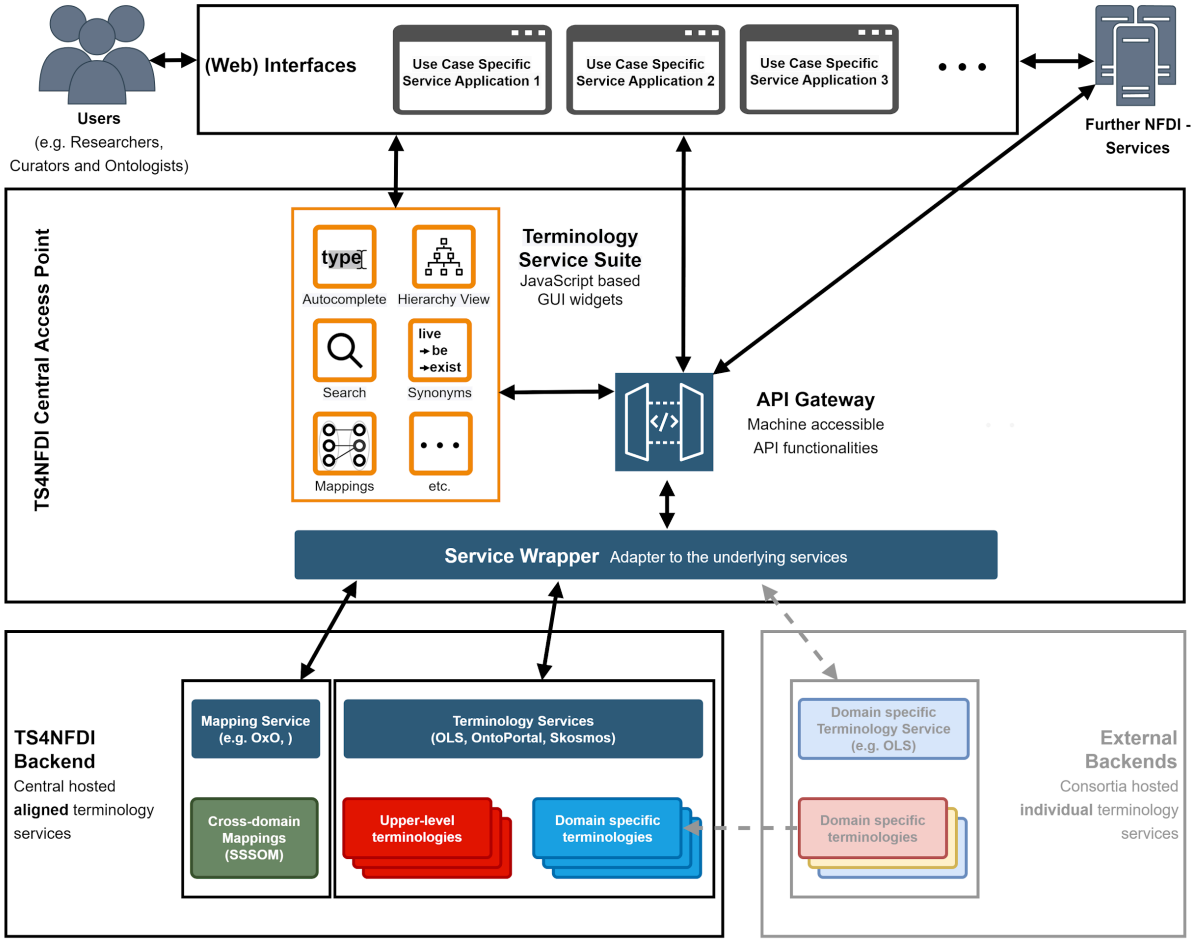

Figure 1: Overview of the TS4NFDI Architecture.

Besides connecting existing solutions TS4NFDI can provide community-defined terminology collections and customised frontends based on GUI widgets for consortia that do not yet have their own offerings (e.g. for BERD@NFDI, NFDI4Bioimage).

The proposed TS4NFDI service can potentially be integrated into multiple (planned) NFDI basic services (e.g. Data Integration Service, Data Management Plans, Knowledge Graph Infrastructures, Electronic Lab Notebooks) as a service backbone to semantically annotate digital artefacts.

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

The underlying technologies and open source solutions OLS, OxO, OntoPortal and Skosmos are in operation in a number of TSs for years (i.e. OLS EBI [4], BioPortal [6], AGROVOC [17] or TheSoz [18]) and are classified as TRL 9.[2] They are actively maintained, stable and publicly accessible. These solutions are applied and enhanced by several NFDI consortia (NFDI4BioDiversity, NFDI4Chem, NFDI4Health, NFDI4Ing) to the needs of the specific NFDI community. These adapted and extended services are operational in the NFDI infrastructure and in federation with other services for a time period of up to 12 months and can be classified as TRL 7 - 8 with regards to OLS systems with separate backend and frontend components (NFDI4Chem, NFDI4Health, NFDI4Ing). In addition the TS of NFDI4Health already provides an API gateway as a starting point for the proposed basic service.

3. SWOT Analysis

Strengths	**Weaknesses**
Multidisciplinary A broad coverage of knowledge from various domains is accessible. There is a broad demand and different approaches in the context of terminology services are already in place/in development. Broad knowledge and competence within the different consortia is available and constructive feedback can be acquired. This benefits not only technical solutions but also consulting services and workshops. Collection of requirements from multiple domains is available A comprehensive overview of all consortia was produced to assemble what services are developed or needed. Due to the broad demand and the many approaches in terms of terminologies and services within various consortia an extensive set of requirements is available. --- Resources can better be shared There are a number of different services supported by the domain NFDIs but no NFDI consortium itself has enough resources and expertise for the implementation of all necessary features and	Coordination with too many stakeholders necessary There is no strategy or architecture for a central system in place yet and we cannot build on an independent system since different solutions are currently supported by the different consortia. *Mitigation: start with three partners, and a maximum of three different TS systems; in addition, TS4NFDI will be in close exchange with the Section Working Group, integrating requirements from all consortia and setting up different use cases.* --- Siloed, independent solutions and developments Systems differ in their maturity and development of functionalities and cannot easily be harmonised and integrated.

[2] https://ec.europa.eu/research/participants/data/ref/h2020/wp/2014_2015/annexes/h2020-wp1415-annex-g-trl_en.pdf

services for efficient cross-domain terminology use. The robustness and security as well as the scalability of the technology will improve through joined forces and resources in TS4NFDI. Additionally, consortia which do not have a service planned in their proposal can use one of the solutions provided by the commonly guided solutions offered by TS4NFDI. This strength also includes terminology and service resources. Data redundancies can be reduced by a standardised interface and reliability of the systems can be ensured. Deployed MVPs/ productive systems available A number of terminology service solutions have been or are being developed and deployed within different NFDI consortia based on open source solutions such as OLS, OntoPortal and Skosmos. Those services provide comprehensive functions to build upon and a large user community. Facilitates the overall management of interoperable terminologies Using shared and standardised technical services reduces interoperability issues and supports harmonisation activities and other services of different working groups.	*Mitigation: (i) a comprehensive analysis of existing terminology services, (ii) early involvement of all persons responsible for NFDI TS as consultants in the initiation phase and closer involvement in subsequent phases, (iii) sorting of requirements according to relevance and (iv) early pilot implementation of a central system will give us a high chance to integrate all different NFDI TS* Bias to terminology needs of main contributors The participating institutions and their terminology services are all located in the sciences but terminology needs of humanities differ from existing use cases. *Mitigation: consortia from the humanities are consulted and at least one institution from these consortia will be a participant in the integration phase of TS4NFDI.* Lack of resources to implement relevant features *Mitigation: if TS4NFDI identifies requirements and commonly used tools very early for all consortia, it should be possible to use development resources more efficiently. Nevertheless, if TS4NFDI identifies major development gaps which cannot be filled using TS4NFDI funding alone, TS4NFDI will identify additional funding opportunities.*
Opportunities Central NFDI entry point to terminologies A consolidated central entry point that bundles the needs of all NFDI consortia makes it easier to set up a common infrastructure for the whole NFDI and to adhere towards the FAIR criteria. --- High demand on cross-domain TS solutions beyond NFDI The need for cross-domain semantic services is also present internationally, independent of NFDI. We have competencies in NFDI for all TS systems and plan to integrate them. This will allow us to significantly participate in EOSC developments.	**Threats** A central terminology service does not fit into the NFDI infrastructure Currently, we have no common architecture of NFDI and might develop an architecture that the central terminology service does not fit into. *Mitigation: early communication and exchange within the sections and with the other basic services. TS4NFDI will intensively exchange with the section Common Infrastructure to explore how the central service would fit in.* --- Incompatibility with EOSC Currently, it is unknown what kind of services might be integrated within EOSC. They might be incompatible with the solution proposed in TS4NFDI. *Mitigation: early communication and exchange with EOSC/ELIXIR developers.*

Table 3: SWOT Analysis

4. Working Concept for the Development of the Basic Service

The overarching vision of TS4NFDI is achieved by addressing a number of key objectives:

Key objective 1: Provision of a consortia-agreed, centralised TS architecture to align the various services on an infrastructure level.

Key objective 2: Provision of a TS-Suite for uniform terminology access to support terminology provision, management, curation, publication, archiving and subscription.

Key objective 3: Foster consens-finding, harmonisation and alignment of terminologies across disciplines by integrating TS4NFDI service components and adopting common practices

In order to realise the vision and key objectives of TS4NFDI, the requirements of the various stakeholders and existing services will continue to be gathered in cooperation with the NFDI consortia, the sections and Base4NFDI. Based on the experience of the NFDI so far, we expect a heterogeneous landscape of terminologies, management tools and terminology services across the disciplines. From the results, an alignment process will be derived to develop a centralised TS architecture agreed upon by the consortia to enable consistent terminology provision, access, curation and archiving of terminologies.

The TS4NFDI basic service aims to extend the scope of the existing TSs beyond provision and access. The TS-Suite will offer various functionalities for a holistic ontology management and consensus-finding across disciplines including terminology curation, version management and integration of (cross-domain) terminology mappings. It will apply best practices and tools for assessing and improving the quality and validity of terminologies. This will include in particular the processing of quality metrics such as consistency, completeness, conciseness, and usability of terminologies. These parts of the developments will be carried out in close cooperation with the working group Ontology Harmonization and Mapping and further stakeholder in the consortia working on ontology development and curation.

4.1 Service initialisation concept

The **service initialisation strategy** includes **first** a comprehensive overview of all consortia: how are terminologies used and applied, what services are being developed or required, what goals consortia have formulated and what's the current state of progress. All requirements for domain-specific services as well as for an overarching service will be collected and prioritised through the organisation of workshops and surveys. The initialisation phase will include an in-depth analysis of existing software and APIs used for terminology management and provision. European TS developments and requirements from infrastructure initiatives such as

the European Open Science Cloud (EOSC) or ELIXIR will also be considered in this analysis to ensure the verification of compatibility with national and international developments.
Second, use cases relevant to the NFDI consortia and NFDI as a whole will be identified, which will be used to evaluate the pilot services and enhanced functionality developed in the initiation phase. **Third**, a detailed architecture will be developed (see Figure 1) for the NFDI TSs and workflows based on the use cases will be specified. In addition, TS4NFDI will implement a first pilot where three different TS platforms (OLS, Skosmos and OntoPortal) will be integrated into an aligned TS backend. The API gateway developed by NFDI4Health will be used as a starting point for the TS4NFDI Central Access Point and service wrappers will be developed to align the different backends. Furthermore, a mapping service (e.g. OxO) will be implemented into the backend to support cross-domain mappings applying SSSOM standard. **Fourth**, at the end of the initialisation phase, TS4NFDI will provide reusable and modular GUI widgets that will be available for any consortium to either integrate into existing TSs or for setting up a domain-specific TS web interface. **Fifth**, the functionality of both the TS-Suite and the centralised TS pilot will be evaluated based on the use cases identified in the requirement analysis.

4.2 Development and integration outlook

After a successful initialisation phase and proof of concept of the architecture and the resulting pilots, the service integration strategy will mainly focus on the integration of the central TS into an overall NFDI architecture by 1) providing access to the central TS components via web interface, 2) the integration of further domain-specific NFDI TSs at the backend level, 3) support and training by presentations and documentation, 4) the implementation of advanced functionalities for terminology use and mapping not covered in the initialisation phase based on the needs of the consortia and 5) the development of a process for updating terminologies in the TS4NFDI backend automatically.
In addition, to achieve service integration, development groups will be formed from the NFDI consortia to carry out load and functional testing. Direct feedback from the development groups will ensure that the developments do not miss the needs of the different consortia.
The result of such an integration phase would be a first centralised system that integrates a significant number of existing services and has passed initial load and functional tests.

4.3 Ramping up for Operation

To scale the TS4NFDI service to a larger user group, several requirements need to be addressed: Since the OLS-based TSs from the PIs NFDI4Chem and ZB MED are actively

maintained, stable and operational for years at TIB and ZB MED, respectively, TIB and ZB MED could provide the necessary hardware for the TS4NFDI backend.. This implies that no additional hardware is required since the applications will be hosted within the consortia themselves. To maintain the TS4NFDI service effectively, a staff dedicated to updating the software and terminologies is required. This includes staying in contact with the consortia to understand their specific needs and incorporating relevant updates. Second, a monitoring system should be in place to ensure the smooth operation of both the central service and the individual services provided by the consortia. Third, personnel would be needed to handle day-to-day maintenance tasks, such as troubleshooting, bug fixing, and ensuring overall service availability. Fourth, the TS4NFDI service should be connected to the European Open Science Cloud (EOSC) to enable integration and interoperability with other research infrastructure and services across Europe.

4.4 Risks and Challenges

Since other services may depend on the TS4NFDI service, ensuring its availability is crucial. The risk of service downtime or failures can disrupt the operations of dependent services and negatively impact researchers and users. To mitigate this risk and to ensure continuous service availability, setting up a backup system can provide redundancy and enable a smooth transition in case the primary system fails. Additionally, implementing robust monitoring mechanisms can help detect and address any issues. The presence of deprecated terminologies within the service poses another challenge. To mitigate this risk, close connection with the consortia should be maintained and regular checks should be conducted to ensure that the terminologies used in the service are up-to-date. This includes monitoring and verifying if the source terminologies have been deprecated or if their location has changed.

5. Work Programme

5.1 Overview of work packages

The TS4NFDI initialisation phase work program encompasses four work packages. The first one will focus on an in-depth requirement analysis, including an exhaustive review of available technologies and tools in the NFDI context, as well as an overview of consortia specific needs in regard to terminology use and management. WP1 will work in close collaboration with the consortia and their stakeholders in order to identify the most relevant use case scenarios, review and document them to serve as basis for the other work packages. In WP2, we will concentrate on designing the system's overall architecture and use cases workflows, specifying interfaces based on the requirements collected in the previous work package and collecting

information necessary to their implementation. Additionally, this work package will be responsible for the coordination and harmonisation of different consortia developments by documenting all functionalities of software and tools that can be reused across consortia. Based on those specifications, WP3 will be developing a pilot system connecting an initial set of backend services as well as a ready to deploy suite of tools and software to be made available and reused for an easy adoption in all consortia. In the last package, we will work on the use case implementation and testing of all necessary functions and their evaluation in terms of usability as well as the elaboration of the integration phase based on the results of the use cases feasibility study.

Work package	Deliverables (D) and milestones (M)	Responsible partner
WP 1	**D1.1** Report on terminology services and technologies (Month 4) **D1.2** Document(s) with requirements from different perspectives (Month 4) **D1.3** Document(s) with Use Cases (Month 5) **M1.1** Network of terminology service experts and contacts across NFDI consortia, NFDI sections, workgroups and Service Stewards established (Month 6) **M1.2** Ranked list of required functions and Use Cases are agreed upon with consortia via the section and its working group(s) (Month 6)	TIB (lead) InfAI ZB MED
WP 2	**D2.1** Technical document on architecture and service functions is available (Month 6 first version, iterative updates till Month 12) **D2.2** Workflow specification for use cases (Month 3 first specifications ready, iterative extensions till Month 12) **D2.3** API specification for a selected set of functions (Month 3 first specifications ready, iterative extensions till Month 12) **M2.1** Technical documents for the basic service are released (Month 12)	InfAI (lead) TIB ZB MED
WP3	**D3.1** API gateway is available (V1 Month 6, iterative updates till Month 12) **D3.2** Pilot implementation of a centralised system with at least three different backend databases (Month 10 first version, iterative updates till Month 12) **D3.3** Readiness to deploy a TS based on the TS-Suite V1 (Month 12) **M3.1** Pilot centralised service up and running (Month 12)	ZB MED (lead) TIB InfAI
WP 4	**D4.1** Demonstrator(s) with use cases implemented (Month 10 first version, iterative updates till Month 12) **D4.2** Report on evaluation of uses cases available (Month 12) **D4.3** Recommendations for the implementation phase available (Month 12)	TIB (lead) InfAI ZB MED

Table 4: Work programme with work packages, deliverables, milestones, and responsible partners.

5.2. Detailed work programme

5.2.1 WP1 “Technological review of terminology services and collection of requirements” (Month 1-6)

The goal of this work package is to conduct a technical review of existing software solutions, collect requirements from all consortia and their communities, service providers and further stakeholders, and formulate use cases based on these requirements. These are critical steps in the development of any basic service, as it will help to ensure that the service meets the needs and expectations of most consortia and is technically feasible. WP 1 cooperates closely with the Base4NFDI service stewards, the respective NFDI sections and their working groups. It will identify experts with regards to terminologies, the application of these and also with service providers using terminologies within the NFDI consortia.

Within WP1, we consolidate and complete the technology review of the TS landscape in the NFDI, EOSC and the scientific communities initiated by the working groups of the NFDI section Metadata to include all 26 consortia of the NFDI. The review will also cover collecting information on the management, curation and development of terminologies to create a more general overview of software and tools used for these tasks. The collection is completed by gathering information about further NFDI services that want to use TSs and their content via APIs or widgets. For the technical review of applied solutions or those in development technology stacks, functionalities, standards, formats and APIs are documented. The review will be organised by surveys across the consortia and available technical and other documentations.

Secondly, WP1 collects, analyses and documents requirements expressed by the consortia and their communities towards TSs. The requirement analysis process will be done in close cooperation and with support of the Base4NFDI service stewards using the previously founded network with the consortia. Requirements will be collected via surveys, focus groups or interviews. Methods may vary depending on available resources. Collected requirements will be reviewed, stakeholders, actors, common issues and patterns identified and grouped. As a start the work package will build on the collected user requirements and user stories from DataPlant, NFDI4Biodiversity, NFDI4Chem, NFDI4Culture, NFDI4Health and NFDI4Ing and their respective activities for their scientific communities. Requirements are expected to fall into categories like (cross-domain) lookup functionalities, mapping between ontologies, connecting domain specific services to the terminology service, archiving, storage, publication, visualisation, curation and development of terminologies. Based on the requirement analysis process, an evaluation and prioritisation will be made in cooperation with the consortia to

identify the most prominent use case scenarios to accompany the initialisation phase and possible transfer into the integration phase.
The use cases will comprise actions and interaction between actors and the terminology service describing features, functions and workflows. Use cases will be documented, reviewed and refined. They will be addressed during the design, realisation, testing and evaluation steps in the following work packages.

5.2.2 WP2 “Architecture of centralised system” (Month 1-12)

The main focus of this work package is the design of the TS4NFDI centralised system architecture, the specification of WP1 compiled workflows supporting terminology development, curation and mapping, the coordination and harmonisation of schemas, content and tools necessary for the implementation of collected requirements and specified use cases workflows.
The architecture (Fig. 1) of the centralised system will encompass a central API gateway that acts as a single access point for API requests over various NFDI TSs. The set of API endpoints will be specified based on use cases requirements for a cross-domain TS access. A service wrapper will be specified for the existing TSs, it will act as an adapter to the underlying services based on a specified mapping. The gateway is responsible for routing incoming requests from the API gateway to the service wrapper and back. The service wrapper is responsible for any necessary format translations, protocols, or additional business logic to connect to any of the underlying backend services. This can be realised using a microservices framework. With this approach it would be possible to interact with external backends which provide the same API as the central hosted backend. In addition to the upper-level and domain-specific terminologies, the TS4NFDI backend will also include (cross-domain) mappings which are based on SSSOM. A first version of the centralised TS components will be released early. In close cooperation with WP3, iterative revisions and improvements of the proposed architecture specification will be released.
Based on the use cases requirements collected in WP1, necessary workflows for terminology management will be specified and documented with the Business Process Model and Notation (BPMN). Furthermore, the frontend components are specified based on the use cases and requirement analysis. These components will be developed in WP3 and can then be used like a modular system to flexibly assemble a frontend according to the desired functions. The workflows for the use cases and the corresponding components are used in WP4 to evaluate the functionality of the resulting TS systems.

5.2.3 WP3 “Pilot integration of centralised system and proof of concepts” (Month 1-12)

The goal of WP3 is proof of concept for (i) a centralised system with an API gateway to enable harmonisation of different terminology service APIs and (ii) a TS-Suite to provide a number of modular frontend functionalities which can be used to either use the already existing services or to set up a completely new TS by using any of the integrated terminology services as backend. WP3 interacts closely with WP1 to address the requirements and use cases expressed by the consortia and their communities towards terminology services. The pilot integration of a centralised system and a TS-Suite will be based on the architecture and use case workflows specified in WP2.

The development of a pilot system starts at the beginning of the initialisation phase. In close collaboration with WP1 and WP2 and based on already existing work of the PIs, different backend systems will be selected for setting up an API gateway and implementing basic API functions for a centralised system. The services will be first set up in a small-scale test environment and later scaled up to avoid complexity issues and to be able to focus on the API functions. OLS, OntoPortal and Skosmos are considered as test systems for the pilot system.

Based on the architecture and interface specifications provided by WP2, service wrappers will be implemented for each of the test services to forward requests to the underlying terminology backend systems and send results back to the API gateway. A modular approach will reduce the effort to integrate new additional systems to the architecture and thereby work as a central entry point for different TSs of the consortia.

Secondly, a TS-Suite with modular frontend functionalities will be implemented based on the evaluated and prioritised requirements gathered in WP1. They can also include JavaScript based widgets to provide (additional) functionalities and allow the integration into 3rd party (terminology) services. Therefore, existing TSs can be replenished and made available as a ready to deploy set. Accessibility to the TS-Suite will be provided by a web application with a graphical user interface.

5.2.4 WP4 “Application and evaluation of use cases driven services” (Month 6-12)

In WP4, the developed pilot of the centralised system and available service components are evaluated against the use cases derived from the user requirements of WP1 to assess the functionality, effectiveness and efficiency of the architecture, interaction of the services and components and their implementation. Demonstrators and prototypes are provided using respective service components to test and evaluate use case driven services. The evaluation

will be conducted using both qualitative and quantitative methods, such as user testing, heuristic evaluation, and performance testing by (i) TS4NFDI co-applicants in correspondence to user requirements and (ii) the wider user community represented by members of the working group Terminology Service of the NFDI section metadata.

The service evaluation and user research will provide valuable feedback to all other WPs and will thus be of high relevance for the next iterations of the design and development of the TS4NFDI services. The evaluation report will provide a detailed analysis of the pilot system performance, its compliance with the use cases and recommendations for the implementation phase.

6. Required Support Actions from Base4NFDI / NFDI Sections / NFDI Consortia

Support from	Work package	Contact Person Basic Service
Base4NFDI	WP1: Support in setting up questionnaire for requirement analysis	Oliver Koepler
Section Metadata, WG Terminology Services, WG Ontology Harmonisation and Mapping	WP1,3: Support in ranking of functionalities to be implement	Roman Baum
Section Metadata	WP1,2: Support in selection of use cases, discussion of workflows	Oliver Koepler, Naouel Karam
Section Infrastructure	WP2: Architecture discussion	Naouel Karam

Table 9: Support request

III Appendix

a) Bibliography and list of references

[1] A. Ledl, ‘Das BAsel Register of Thesauri, Ontologies & Classifications (BARTOC)’, 2015, Accessed: Jan. 23, 2023. [Online]. Available: https://opus4.kobv.de/opus4-bib-info/frontdoor/index/index/docId/1677

[2] J. Voß, J. M. Agne, U. Balakrishnan, and M. Akter, ‘Terminology Registries and Services’, Nov. 28, 2016. doi: 10.5281/zenodo.166717.

[3] S.-Y. Yang, ‘OntoPortal: An ontology-supported portal architecture with linguistically enhanced and focused crawler technologies’, *Expert Syst. Appl.*, vol. 36, pp. 10148–10157, Aug. 2009, doi: 10.1016/j.eswa.2009.01.004.

[4] S. Jupp, T. Burdett, C. Leroy, and H. Parkinson, ‘A new Ontology Lookup Service at EMBL-EBI’, in *SWAT4LS*, 2015.

[5] ‘Welcome to the OntoPortal Alliance’, *Ontoportal Alliance*. https://ontoportal.org/ (accessed May 15, 2023).

[6] N. Noy *et al.*, ‘BioPortal: A Web Repository for Biomedical Ontologies and Data Resources.’, Jan. 2008.

[7] S. Jupp, T. Liener, S. Sarntivijai, O. Vrousgou, T. Burdett, and H. Parkinson, ‘OxO - A Gravy of Ontology Mapping Extracts’, presented at the International Conference on Biomedical Ontology, 2017. Accessed: Jan. 11, 2023. [Online]. Available: https://www.semanticscholar.org/paper/OxO-A-Gravy-of-Ontology-Mapping-Extracts-Jupp-Liener/98ffd29ab7b3d454c97dea76692bd4f6f0610623

[8] O. Suominen *et al.*, ‘Publishing SKOS vocabularies with Skosmos’.

[9] ‘SKOS Simple Knowledge Organization System Reference’. https://www.w3.org/TR/skos-reference/ (accessed Jan. 23, 2023).

[10] ‘Dokumentation der Schnittstellen zu DANTE’. https://api.dante.gbv.de/ (accessed May 15, 2023).

[11] P.-Y. Vandenbussche, G. Atemezing, M. Poveda-Villalón, and B. Vatant, ‘Linked Open Vocabularies (LOV): A gateway to reusable semantic vocabularies on the Web’, *Semantic Web*, vol. 8, pp. 437–452, Jan. 2017, doi: 10.3233/SW-160213.

[12] ‘Gemet’. https://www.eionet.europa.eu/gemet/en/themes/ (accessed May 15, 2023).

[13] ‘GND - Home’. https://gnd.network/Webs/gnd/DE/Home/home_node.html (accessed May 15, 2023).

[14] ‘GeoNames’. https://www.geonames.org/ (accessed May 15, 2023).

[15] DataPLANT Team, ‘Swate’. https://github.com/nfdi4plants/Swate (accessed Jan. 23, 2023).

[16] N. Matentzoglu *et al.*, ‘A Simple Standard for Sharing Ontological Mappings (SSSOM)’, *Database*, vol. 2022, p. baac035, Jan. 2022, doi: 10.1093/database/baac035.

[17] I. Subirats-Coll *et al.*, ‘AGROVOC: The linked data concept hub for food and agriculture’, *Comput. Electron. Agric.*, vol. 196, p. 105965, May 2022, doi: 10.1016/j.compag.2020.105965.

[18] B. Zapilko, J. Schaible, P. Mayr, and B. Mathiak, ‘TheSoz: A SKOS representation of the thesaurus for the social sciences’, *Semantic Web*, vol. 4, no. 3, pp. 257–263, Jan. 2013, doi: 10.3233/SW-2012-0081.

b) Gantt Chart

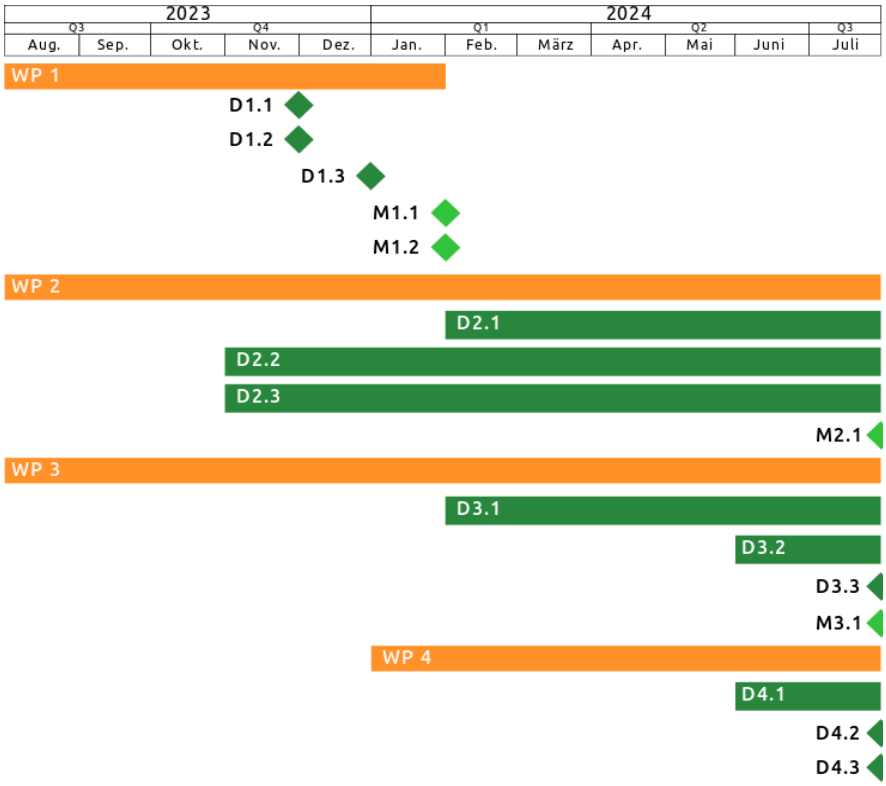

Figure 2: Gantt chart of the overall work programme with work packages, deliverables and milestones