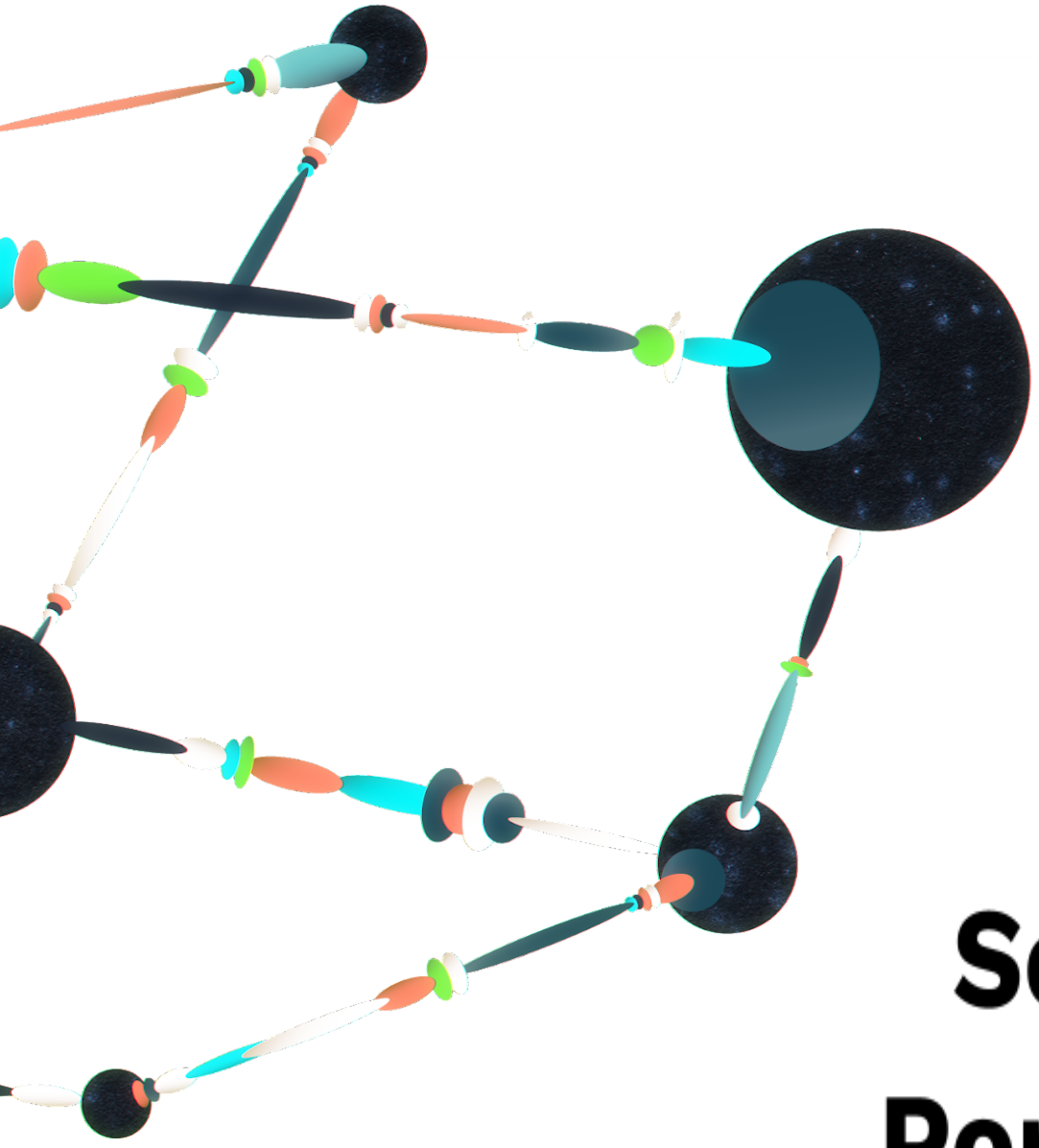

Service Portfolio Report

2023

NFDI 4
BIODIVERSITY
BIODIVERSITY, ECOLOGY & ENVIRONMENTAL DATA

Authors

Maria Meister[1], Manuel Feser[2,] Lena Steilen[3], Birgitta König-Ries[4], Bernhard Seeger[5], Barbara Ebert[1], Frank Oliver Glöckner[1]

[1] German Federation for Biological Data (GFBio e.V.), Mary-Somerville-Str. 2-4, 28359 Bremen, Germany
[2] Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Corrensstraße 3, 06466 Seeland, Germany
[3] Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen, Burckhardtweg 4, 37077 Göttingen, Germany
[4] Friedrich-Schiller-Universität Jena, Leutragraben 1, 07743 Jena, Germany
[5] Philipps-Universität Marburg, Hans-Meerwein-Straße 6, 35032 Marburg, Germany

Published by

German Federation for Biological Data (GFBio e.V.), Mary-Somerville-Str. 2-4, 28359 Bremen, Germany

We would like to thank each and every one of our colleagues within NFDI4Biodiversity and especially the core team of Task Area 3, the members of the SIG2 Service and all partners, for their valuable contributions. They supported the planning and organisation of the Service Portfolio and Service Quality Strategy as well as the documentation and delivery of the services, making the efforts on NFDI4Biodiversity service monitoring a story of success.

Thanks also to the partners and collaborators beyond NFDI4Biodiversity for their valuable support and shared vision.

The work was supported by the German Research Foundation DFG under the grant agreement number 442032008 (NFDI4Biodiversity). NFDI4Biodiversity is part of NFDI, the National Research Data Infrastructure in Germany (www.nfdi.de).

Table of Content

Executive Summary

NFDI4Biodiversity serves a broad stakeholder group in biodiversity research and conservation with a wide variety of needs with respect to handling biodiversity and environmental data. To meet these needs, the consortium partners provide a broad Service Portfolio and aim to develop it further in close cooperation with users from research and practice.

Over the past three years, NFDI4Biodiversity has collected information for a basic Service Portfolio, and designed a strategy for quality assurance and quality control of partner services. Several community services were developed or significantly improved. Furthermore the project has fostered a common understanding of resource provider structures and needs.

- **Service Portfolio:** As a basis for eventually structuring and consolidating tools and technologies used in the community, 80 services from 23 partners have been catalogued. Several community services have been developed or significantly enhanced through NFDI4Biodiversity project funding. The GFBio Data Centers in NFDI4Biodiversity provide an archiving backbone and have a special role as both Service and Data Providers.
- **Helpdesk** has been established which is collaboratively organised by partner organisations as a central point of access for user support.
- **Quality of services:** Establishing quality assessments within the consortium is crucial to offer dependable and high-quality services. For this aim a quality management strategy has been proposed which includes a primary quality as well as developmental goals, specific quality objectives, and three service areas with tailored quality assurance processes. In addition, service quality monitoring relies on Key Performance Indicators (KPIs) as critical benchmarks for evaluating service performance. This strategy aligns with the consortium's vision of being a reliable and transparent infrastructure provider.
- **Promoting the FAIR data principles:** The consortium promotes the implementation of the FAIR data principles for scientific data management and stewardship through community and partner projects. NFDI4Biodiversity services facilitate working with and creating FAIR data. At the community level, partners collaborate on implementation of Bioschemas.org to improve findability on the web and the ABCD consensus elements initiated in the GFBio project in order to increase interoperability. A new ontology service was launched based on technology of the OntoPortal Alliance. At the partner level, best practices were established to mobilise species occurrence data, improving accessibility, interoperability, and reusability.
- **Sustainable portfolio management structures:** Measures are in place to guide service providers in the consortium towards implementation of common standards and technologies. A sustainability assessment has been

started to investigate services provided in the consortium in terms of costs and models, considering institutions' working conditions.
- **Providing a reference implementation for the Research Data Commons (RDC):** Progress has been made to realise the RDC as a central self-service platform for data and services. Components on the storage and mediation layer are being tested through use cases and RDC pilot applications to showcase the benefits of using its tools. Negotiations have been started in the NFDI Section Common Infrastructure of the NFDI for uptake of the architecture and collaboration across consortia.
- **Policies:** Resource Provider Workshops were held to define requirements for developing policies for data and service provider roles. Focus was the legal framework, especially regarding licences, as well as guidance for providing FAIR resources and certification for CoreTrustSeal (CTS).

1 Introduction

In order to understand various phenomena related to species and ecosystem diversity and function researchers need access to a wide range of biodiversity and environmental data. This data comes from a wide variety of sources that are often only temporarily accessible or are difficult to find. In areas like ecosystem research, ecology, taxonomy, bioinformatics or biodiversity monitoring, many standards and services have been developed to improve data and service availability and to provide software-based solutions to both users and providers of resources. Nevertheless, the landscape of data and service providers and users is still rather fragmented.

NFDI4Biodiversity [1] is a consortium under the umbrella of the German National Research Data Infrastructure (NFDI) [2] with about 50 partners, ranging from research organisations and scientific IT centres to expert organisations, government agencies and citizen science initiatives. It represents stakeholders from different national and international data initiatives. A central activity is the development of a coherent service portfolio to mobilise, store and process data to make them available to interested user groups. The community of interest includes stakeholders from science, politics, nature conservation and landscape management. Their needs are the guiding basis for our activities. For this purpose, datasets are prepared in accordance with the FAIR principles to better secure, organise, publish, and reuse valuable data for a variety of purposes in science and society.

This report describes the structure and the progress of the Service Portfolio development in the first 30 months of the NFDI4Biodiversity project and the activities that are planned for the future.

1.1 Explanation of Terms and Structures

This chapter provides an overview of specific terms used in the report and the relevant structures in the consortium the report refers to.

Task Areas (TA)

The NFDI4Biodiversity work program is structured in five major work packages, known as Task Areas (TA), with multiple measures (see Figure 1).

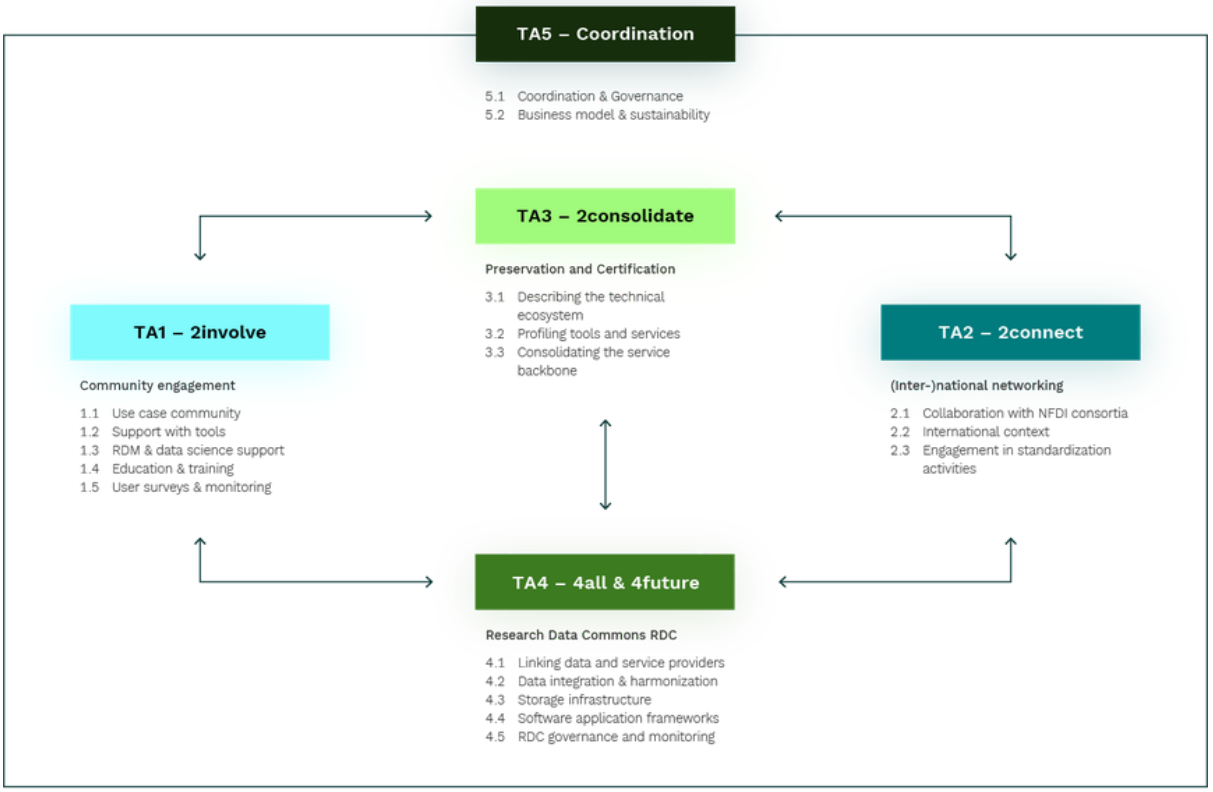

Figure 1: Overview of the Task Areas and the corresponding measures until 2025

Task Areas 1 to 4 are centred around technical and operational aspects, involving activities like coordinating expert participation and building international networks. Task Area 5 addresses the organisational necessities and guarantees the enduring availability of the developments made within NFDI4Biodiversity even after the project's conclusion.

Measure (M)

Task areas are divided into measures, structured along milestones to develop deliverables. The names used throughout the report include the Task Area as prefix. For example, Measure 1 in Task Area 3 is spelled TA3M1.

Special Interest Groups (SIG)

Special Interest Groups (SIG) are implemented for cross-cutting topics within the consortium. SIGs are the think-tanks of NFDI4BioDiversity and prepare decisions for the Steering Committee. The Special Interest Group 2 Service Monitoring and FAIRness (SIG2 Service) was established to address the topic of quality assurance. It works across Task Areas 1, 3, 4 and 5 to define a basic quality strategy and establish a monitoring system, including key performance indicators

(KPIs) for the usability and FAIRness of all services and archives of NFDI4Biodiversity.

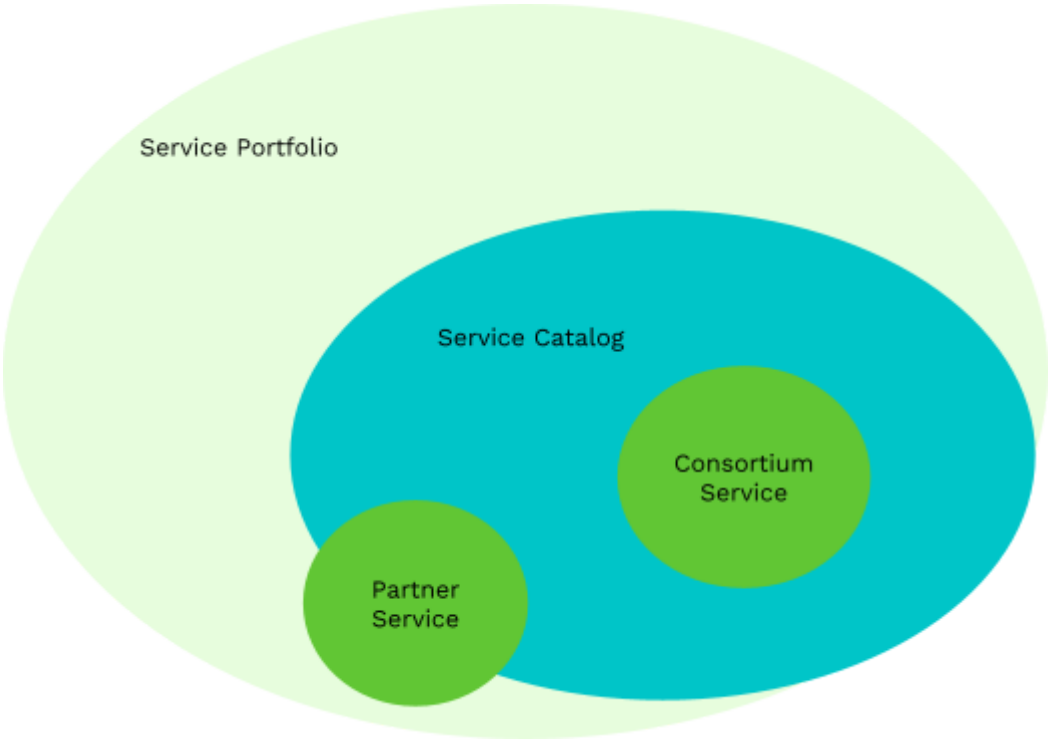

Figure 2: Overview of the Service Terminology

The Service Portfolio is the collection of all services provided within the consortium. The Service Catalog is a subset that will be publicly promoted by the consortium and for which certain criteria need to be fulfilled. Services can either be Consortium or Partner Services, where Consortium Services fulfil more quality criteria.

Service

We use the definition developed across NFDI consortia, according to which a service is understood as “a technical-organisational solution, which typically includes storage and computing services, software, processes, and workflows, as well as the necessary personnel support for different service desks'' [3]. NFDI4Biodiversity services can be hosted on the reference implementation of the Research Data Commons (RDC) or be provided directly by a partner organisation. They are categorised according to the schema provided in Chapter 2, Table 1. For portfolio development and quality management purposes, we distinguish two groups of services (see Chapter 5):

- **Consortium Services** have been assessed in a structured process and fulfil a defined set of quality criteria, including support through the consortium´s central Helpdesk.
- **Partner Services** are services offered by a partner organisation of NFDI4Biodiversity that are open for use by the Community, such as specialised data portals or APIs to partner resources, complying to a subset of the quality criteria.

Resource / Service Provider

All partners are resource providers in the sense that they provide digital objects or processes, such as data and metadata, publications, software, workflows, services and training materials [4]. Service providers are partners which open their own existing infrastructures and tools along the data life-cycle [5] for the community or develop community solutions within the work programme of NFDI4Biodiversity.

Service Portfolio

The Service Portfolio is a registry of services offered (and intended to be promoted) by the NFDI4Biodiversity partners (see Chapter 2). It serves internal purposes to understand, structure and eventually consolidate tools and technologies used in the community.

Service Profile

When registering a service in the Portfolio, providers fill out a structured Service Profile form (see Appendix B). A Service Profile contains all information about a

service, either needed for the user or for the implementation of the work programme.

Service Catalog

The Service Catalog is a subset of the services in the portfolio (see Figure 2) endorsed and publicly promoted by the consortium. These services will be subject to a comprehensive quality assurance and findable in a Service Discovery Portal.

1.2 Partner Network and Governance

The NFDI4Biodiversity consortium consists of about 50 German organisations - from research institutions and scientific IT service centres over government agencies to natural history collections and civil society initiatives. The partners pool their scientific and technical expertise to provide a broad Service Portfolio for handling biodiversity and environmental data and to develop this portfolio further in close cooperation with users from research and practice. The cooperation is governed by a cooperation agreement and a joint work programme, which is monitored by a Steering Committee. All partners are committed to the Guidelines for Safeguarding Good Research Practice [6] and the implementation of the FAIR Guiding Principles for scientific data management and stewardship [7]. At this point in time, a common understanding of structures and prerequisites of the resource providers has been reached, enabling the provision of a coherent Service Portfolio and a Service Catalog building on top of it.

1.3 Community added Value

Within the NFDI, the consortium partners offer added value to the professional community, specifically:

- Access to modern technologies and a comprehensive stock of biodiversity and environmental data;
- Methods and tools for archiving, publishing, searching and analysing data that are suitable for everyday use and have been tried and tested in practice;
- An expert forum for the safe and competent handling of data for diverse and responsible use.

This value proposition was formulated as a mission statement and adopted by the Steering Committee in 2022 [8].

1.4 Core Values

The mission statement of the consortium also formulates the values that guide the collaborative work:

Trust: The consortium partners treat the data and information to which they have access with respect and responsibility. The participants are committed to the Guidelines of Good Research Practice.

Transparency: We make existing data discoverable and reusable in accordance with the internationally recognized FAIR data principles. It is important to us to preserve information on provenance so the origin can be traced back and referenced properly.

User orientation: We develop our services based on real-life application scenarios. We learn from and cooperate with the professional community in order to be able to provide applications and training offers that meet the needs of our users.

Quality: Our roots are in science and research. We use testing and documentation procedures to ensure the high quality of our offerings and apply recognized standards wherever possible.

2 Designing the Service Portfolio

2.1 Basic Concept

In the NFDI4Biodiversity project, the partners are ready to provide tried-and tested services for a wider audience and to develop or implement further community services. Collecting information on the multitude of services available in the network is primarily intended to make the portfolio accessible for structuring and eventually consolidating tools and technologies used in the community (Task Area 3), as well as building a marketing strategy and organising shared responsibilities for valuable resources, if needed. Designing the Service Portfolio is an interplay of various activities in the work program:

- Identification and description of the current and emerging technical ecosystem in biodiversity research (TA3M1)
- Profiling of biodiversity tools, technical services and data centres, adopting them as NFDI-RDC recommended and guiding them towards certification (TA3M2)
- Adaptation, enhancement and consolidation of the NFDI4Biodiversity tool and technical service backbone including data pipelines as service (TA3M3)
- Mobilising and archiving existing and newly generated data via use cases engaging biodiversity communities (TA1M1)

- Supporting tools/platforms for early mobilisation of data and data analysis (TA1M2)
- Software application frameworks (TA4M4)
- Governance and Monitoring (TA4M5)
- Sustainability and business models (TA5M2).

Additional input was provided by the SIG 2 Service regarding information on indicators and quality management. The information collected in the initial Service Portfolio is a common basis for the work in these measures.

2.2 Structure

Services are provided by the partner organisations according to the cooperation agreement. They have i.e. responsibility for the design, quality and sustainability of their individual services but profit from expertise, funding and rights to use other partners´ software for development purposes.

To structure the Service Portfolio, the well-established de.NBI categories have been used [9]. Each service is assigned to exactly one category, as the assignment also serves as the basis for the monitoring of the Key Performance Indicators (KPI) associated with the de.NBI categories. Table 1 shows the categories and definitions used for the assignment, as well as some examples. The extensions marked with * are currently not implemented.

Table 1 - de.NBI service components plus NFDI extensions (marked with a * and separated by the green row) [10]

WEB APPLICATION	Software that is installed on a server and can be used by users via a web page and the internet. E.g. BIIGLE [11]
TOOL/APPLICATION	Software that can be downloaded and executed locally on the users' hardware. E.g. RightField [12]
DATABASE	Software providing large amounts of structured data to the user. Usually, the data can be uploaded, accessed, searched, edited and/or downloaded via a client or web browser. E.g. PANGAEA [13]

WORKFLOW/PIPELINE	Software that combines multiple Tools/Applications. It may be used locally or remotely via the internet. E.g. Metabarcoding Pipeline for Freshwater Samples
LIBRARY/API	Collection of pre-implemented functions for a specific task that can be accessed via a well-defined interface. E.g. BfN Checklist API [14]
SUPPORT/CONSULTING	Service with direct user contact for topics going beyond the support for the other services. E.g. RDM Helpdesk [15]
DATA CURATION*	If not included in Support/Consulting: “The activity of managing and promoting the use of data from their point of creation to ensure that they are fit for contemporary purpose and available for discovery and reuse. For dynamic datasets this may mean continuous enrichment or updating to keep them fit for purpose. Higher levels of curation will also involve links with annotation and with other published materials.” [16]
TRAINING*	Standalone training for self-study can be considered a technical service (usually a web application). Generally speaking, training materials often come in the form of specific measures or tutorials that are attached to a service and that are designed to improve the user’s service experience. Our joint understanding of training as a standalone service, however, is not limited to the above and includes materials designed for education in all fields of research data management.
STORAGE*	Provision of storage space for research data as a service to external users. Access is possible via web protocols.

2.3 Version as of August 2023

2.3.1 Development of the First Version

In the first 30 project months a first version of the Service Portfolio was developed in TA3M1. An initial portfolio was defined, which comprised the Core Services of the German Federation for Biological Data (GFBio) [17]. They are being provided by partners in the consortium and are essential to fulfil the NFDI4Biodiversity's mission (e.g. Data Submission, Data Management Planning, Data Search). The initial portfolio was then completed by an inventory (see Appendix A) of the services each partner described as contribution in the project proposal. Additional services of relevance for the community were identified through analysis of the internal use case interviews conducted in the first project year [18]. Those services fulfil for the most part specific tasks, important to the individual service provider or a specific user community.

Each service was described through a Service Profile (see Appendix B). The template for the profile was developed when the information requirements within the consortium exceeded the initial simple listing of the services in tabular form (with name, service provider and service category). Existing solutions to manage services within service portfolios were examined and the Information Technology Infrastructure Library (ITIL) [19] was identified as a good fit. The present NFDI4Biodiversity Service Profile Template is therefore an adapted version of the structure ITIL uses to catalogue its services. The template can be extended when new requirements arise in the consortium. The conceptual work on service sustainability and business models in TA5M2 may for example require querying the funding structure of the services; or the measure for software application frameworks in Task Area 4 may require to classify the services into technical or user-facing services.

The development of the portfolio can be structured into three phases.

1. Top down landscaping phase: In this phase the initial portfolio was created and described by project staff in Task Area 3 Measure 1.
2. Bottom up phase: Here, partners were asked to check the collected information from phase 1 for correctness and completion and add missing services or mark services to be removed from the first version.
3. Consolidation phase: The status of the portfolio was checked for compliance with the profiles and partners were asked to update the profiles to the current version. This phase overlapped with phase 2.

2.3.2 Current Status

In year 3 of the project, a robust process is available for collecting services of relevance in the NFDI4Biodiversity partner group. A formal onboarding process for new services is being developed and piloted with selected services (see Chapter

3.2). Additionally, criteria for the inclusion in the public Service Catalog (including the provision of training and support) are being drafted, and will be piloted shortly.

The Service Profile Template is developed and tested. The profiles are grouped into the Service Portfolio and deployed on the internal project platform. We communicated a process to the partners how they can add their services to the portfolio and defined a process to assure compliance with the template. As not all profiles have been filled out completely, we are currently in contact with partners to update their service profile information. However, all services are already categorised to allow a statistical evaluation (see Figure 3).

2.3.3 Future Development

In project year 4 and 5 we plan to complete the collection of relevant services. As service profiles become more comprehensive, we can begin to query the Service Portfolio for selected organisational and technical properties of the service landscape. The Service Portfolio also delivers initial information for the public Service Catalog and the Service Discovery Portal (milestone of TA4M4). Services were and will be added as the project produces more outputs, either in the single measures, use cases or flexfunds projects.

2.4 Overview of the Services

Until August 2023, 80 services from 23 partners had been registered in the Service Portfolio (see Appendix A, Table A1). As shown in Figure 3, the largest group are Web Applications (26), followed by Databases (25) and Library/APIs (13). Examples of services in these categories can be found Table 1 and 2.

Figure 4 displays the services by their selected areas of application. Service providers can select multiple of six different areas of application. Those areas are “Planning”, “Harmonization”, “Access/Archival”, “Discover/Exploration”, “Processing/Visualization”, “Support/Training”. Additionally service providers can add further areas by listing those in “Others”. By far most services have been classified as “Discover/Exploration” (54) and “Access/Archival” (51), followed by “Processing/Visualization” (34), “Harmonization” (25) and “Support/Training” (21). The areas “Planning” as well as “Others” were selected eight times.

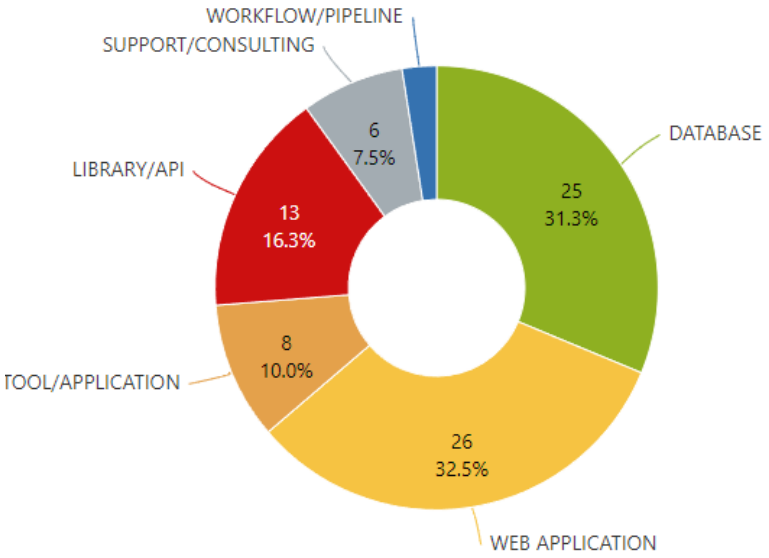

Figure 3: Overview of the NFDI4Biodiversity Service Portfolio. Status: 9th August 2023

NFDI4Biodiversity uses the de.NBI service categories to describe distinct service components, as recommended by the NFDI.

In twelve cases, services were set up de novo or completely renewed (see Table 2). 17 services are supported through the NFDI4Biodiversity Helpdesk (see Chapter 2.5).

The services provided by the partners are partly inhouse solutions or own developments, and partly mature, accepted tools developed in related initiatives, such as GBIF, CETAF, ELIXIR or the Ontoportal Alliance [20].

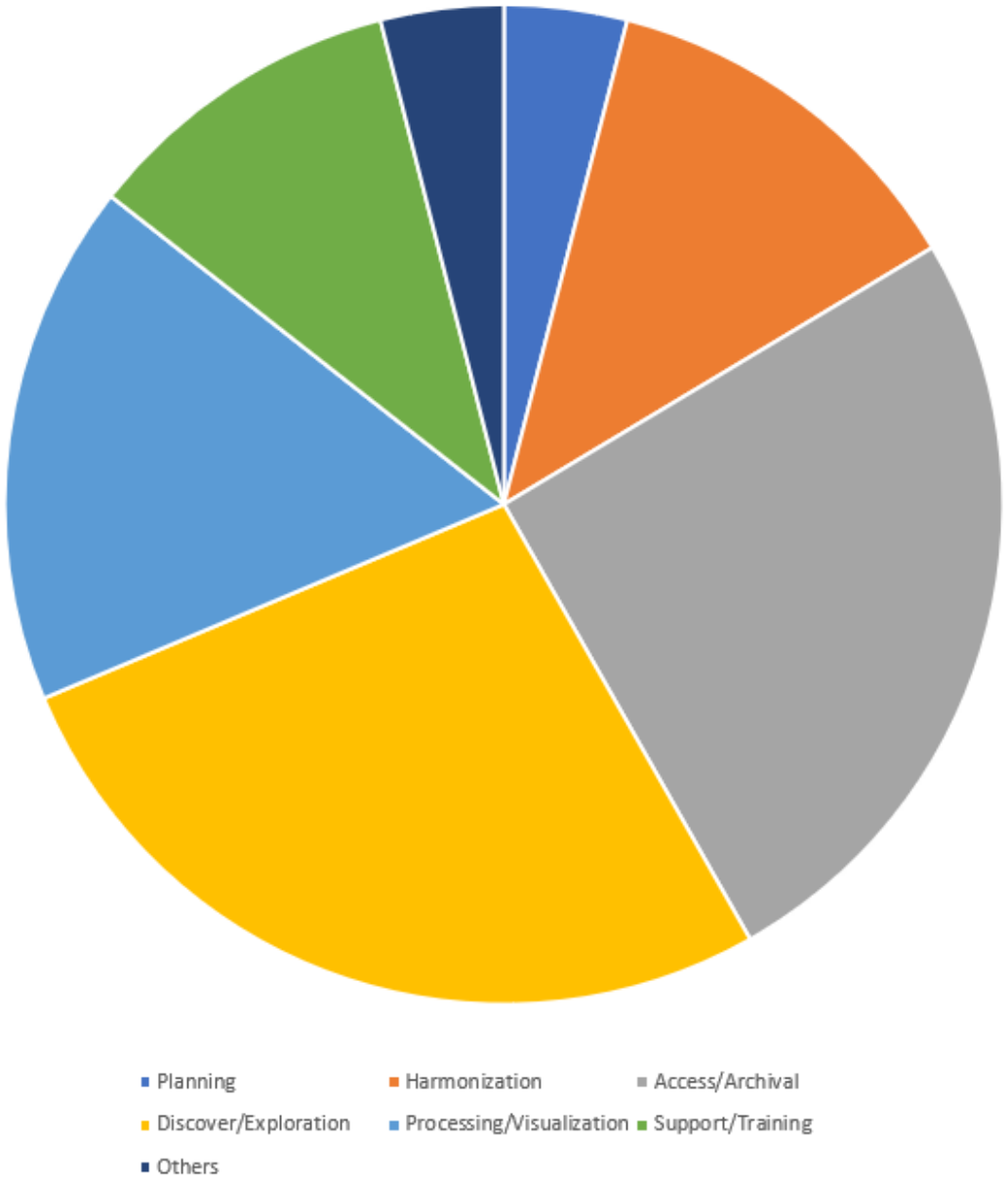

Figure 4: Overview of the NFDI4Biodiversity Service Portfolio by area of application.

Status: 30th August 2023

Information on the provision of services as part of the institutional mission or through project funding is still being completed. In several cases, feedback loops are required for final clarification. In cases where partners open their tried-and-tested infrastructure and services for external users, only the operational staff for Community services (data curators, developers etc.) is project-funded. In many cases, the development of software components is project funded, and then integrated in regular operations. A structured sustainability assessment is ongoing (see Chapter 5.3).

Table 2: Software components developed or upgraded with project funding

Service types and partner roles are named according to the White Paper “interim report reference” provided by NFDI. The references column provides the related project output(s) or a reference providing additional information on the work.

Type	Name	Partner	References
Web application	Atlas of Living Germany (data discovery)	BGBM	[21]
Web application	CRITTERBASE	HIFMB	[22]
Web application	Visualisation, Analysis and Transformation (VAT)	UniMarburg	[23]
Web application	BIIGLE	BIBI	[24]–[26]
API/library	Checklist Editor of the Rote Liste Zentrum	GFBio (commission of API) BfN (operator of Checklist editor)	[14]
API/library	ThünenATLAS	Thünen	[27]
API/library	HydrographR	IGB	[28], [29]
Database	Aruna Core Storage	UniGießen	[30], [31]
Database	BiodivPortal Ontology Service	InfAI	[32], [33]
Database	Scorpion	IPK	[34]
Tool/application	Diversity Workbench Management Software	SNSB	[35]
Tool/application	BEXIS2	UniJena	[36]

The next chapters describe the mode of operation for key infrastructures the consortium partners provide for the community.

A **Helpdesk** is available as a central point of contact for user support, which is collaboratively organised by partner organisations (see Chapter 2.5). The GFBio **Data Centers** in NFDI4Biodiversity are part of the long term data repository and archiving network to be developed as part of TA3M3, they have a special role as both Service and Data Providers (see Chapter 2.6). The **Research Data Commons** (RDC) are the grand technical vision of the consortium; a reference implementation is being developed in Task Area 4 (see Chapter 2.7).

2.5 Central Helpdesk for Tools and Services

Support is provided via a central Helpdesk. The Helpdesk provides first and second level support and is a gateway to the expertise of the whole consortium.

Technically, it is based on a ticket system. The Helpdesk structure and rights management system builds on the GFBio service, which uses Jira (Atlassian), a ticket infrastructure. The Helpdesk service is operated by GFBio e.V. with UniBremen/MARUM as technical provider and staff from partner organisations for first and second level support. Support staff is mainly project-funded.

First level support is given by the Helpdesk Team, which is monitoring the incoming user requests. The Helpdesk Team has access to all requests. If additional support is needed, the Helpdesk Team is activating the second level support by including the respective experts from the consortium to answer specific questions. All NFDI4Biodiversity project members can potentially be added to user requests. Questions targeting a specific service or tool offered by NFDI4Biodiversity partners are directly assigned to the Helpdesk contact defined in the Service Profile (see Appendix A, Table A1).

Within the NFDI4Biodiversity project, a Front Office/Back Office model is being introduced to handle an increasing number of requests around active research data management and data science support with our Helpdesk system (Task Area 1 Measure 3).

The Back Office consists of the Helpdesk Team (first level support) and the expert network in the consortium (second level support). Individual consulting services offered by partners can plug into this system by training staff for Back Office operations.

The Front Office on the other hand is the expert network at the universities and institutions, e.g. the data stewards or data managers that offer on-site support. They can address discipline-specific questions to the Helpdesk Team, or forward the specific questions directly to the Helpdesk Back Office[37].

The Front Office/Back Office model is currently piloted with selected partners, among them the coordination office of a multi-year research initiative on biodiversity research (FEdA [38]). To enhance the Front Office step by step, we are actively networking with the RDM state initiatives (Landesinitiativen), forschungsdaten.info and local Data Stewards in Bremen and Marburg.

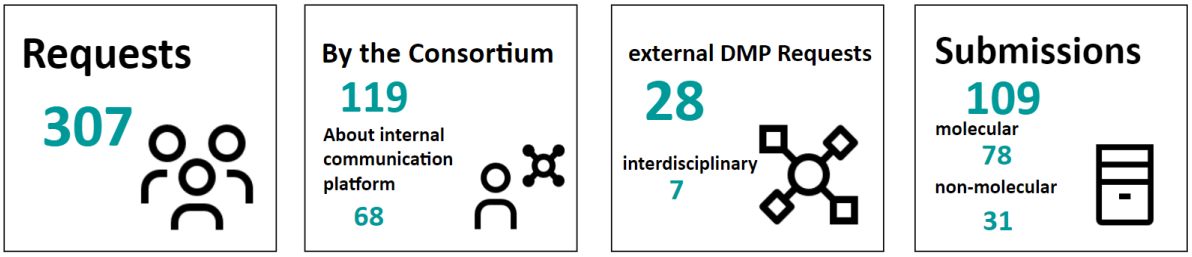

Figure 5: Overview of Helpdesk Requests

Status: August 2023

.

Since the start of NFDI4Biodiversity until August 2023, a total of 307 support requests have reached our Helpdesk (see Figure 5). Of these, 119 requests came from consortium partners mainly regarding the internal communication platform (68), the remaining requests were made concerning practical support with services. The 188 requests from other community stakeholders focused mainly on access, bug reporting and feedback to the provided services, but also included 8 requests for Trainings, 46 requests for data consultations, as well as 4 requests regarding cooperations and 28 requests were made for support with Data Management Plans (DMP). Seven of those were interdisciplinary DMP requests from the FEdA cooperation were supported together with Qualiservice [39], a member of KonsortSWD (NFDI consortium for social sciences). Additionally, 109 total data submissions reached the GFBio Data Centers. A total of 31 non-molecular data submissions and 78 molecular data submissions were recorded.

In addition the Helpdesk provides support with tools, which are missing in the repertoire of the data providers. Service providers who wish to include their service into the NFDI4Biodiversity Service Catalog are requested to provide support through the central helpdesk, i.e. assign a responsibility to staff. The onboarding of BIIGLE and RightField as pilot services (see Chapter 3.2) was accomplished as part of milestone 1.2.1. The services allow the user semantic annotation of their data and support is offered through the central helpdesk. Support workflows have been established as milestone 1.2.2.

2.6 Services provided by the GFBio Data Centers

NFDI4Biodiversity has a strong backbone of domain-specific, institutional data repositories with archives, where data can be stored, published and discovered. The GFBio Data Centers are departments at recognized science institutions dedicated to the management, storage, long term digital archiving and publication of biological and environmental data [40]. They provide (standardised) domain-specific services to the community according to the mission of the organisation.

Out of the 80 services currently registered in the Service Portfolio, more than half are provided by the GFBio Data Centers, ranging from databases with long-term data storage to API and other service categories. The GFBio Data Centers [41] cooperate in a joint submission service and helpdesk as well as a Data portal [42]. The submission service is a brokering framework [43] for distributed archiving and allows researchers to deposit outputs of multi-method projects in dedicated archives, while maintaining relationships between them. A set of consensus documents, tools, data pipelines, standards and technical formats has been agreed for interoperability. Each GFBio Data Center has its own profile [40] and all have decided to demonstrate their trustworthiness in terms of authenticity, integrity, confidentiality, and availability of data and services through aiming the certification with the CoreTrustSeal [44]. The Core Trust Certification was chosen because it has a high level of acceptance in the community (see Chapter 5.2).

Within NFDI4Biodiversity, this group is instrumental to implement the FAIR principles (see Chapter 4). It organises admission of further Data Centers under a joint mandate, to form an organisational network for long-term data storage and archival and close gaps in terms of scope and target communities. Project funding is used to cover additional costs, where curation and management exceed institutional missions.

2.7 Software Components in the Research Data Commons

The Research Data Commons (RDC) provides the architectural framework for a cloud-based research infrastructure designed to support researchers, data providers and data consumers with tools to manage FAIR data products in the field of biodiversity (see BOX 1). The RDC will enable users to reuse heterogeneous data sources, correlate them and perform complex analyses.

RDC includes and enhances pre-existing solutions such as the GFBio Portal and adds new end-user services for data transformation, discovery and analysis as well as a central storage solution. The necessary components are being developed in Task Area 4. An example of a successfully integrated solution is the ontology service [45], which has been enhanced within the project, as well as the Aruna Object Storage (AOS) infrastructure [30], [31], [46], which has been developed by the Justus-Liebig-University Gießen with funding from NFDI4Biodiversity, NFDI4Microbiota and the FAIR DataSpaces project [47]. The ontology service and the search index are already in use, but still require extensive testing until their final release. Task Area 3 aims to support RDC by identifying candidate services for the application layer.

Within NFDI4Biodiversity a reference implementation of the RDC is being developed. Operation and sustainability of a productive RDC are in negotiation with other NFDI consortia. The RDC architecture itself is brought into the NFDI Section Common Infrastructures [48] to discuss joint development as a basic service.

Specifically, we proposed a Data mesh concept [49] as a possible framework for a NFDI-wide architecture. The following chapters highlight the current status of development for the services building the RDC reference implementation.

BOX 1: RDC architecture

The RDC architecture will be structured into four technical layers:

The **Cloud Layer** is the technical backbone based on a multi-cloud infrastructure including for example the de.NBI cloud and GDWG. These clould providers offer scalable functionality for distributed computing as well as cloud storage with near infinite resources such that users are empowered to run compute-intensive jobs or analyze very large data sets in a user-friendly way. In addition, there are cloud services like the Aruna Object Storage (AOS) for managing data in a unified model.

The **Mediation Layer** provides a self-service collection of community-agnostic tools for creating FAIR data products. A community-specific community, for example, a team that works together in one of the Use Case projects of NFDI4Biodiversity, is responsible for a specific data product and also responsible for its development process. A community unterstands the purpose the data product under its responsibility is generally used for, e.g. specific analyses and annotations. In particular, a data product uses common terminology of a certain discipline or subdiscipline. The mediation layer offers tools for data product developers to manage metadata, transform data from a technical data model into a semantic model, and improve data quality. Workflows serve to describe the steps of creating the data product from the source data sets, and thus, it also documents the provenance of a data product.

The **Semantic Layer** provides the community-specific data products that are created and maintained in the Mediation Layer using the technical datasets accessible from the Cloud Layer and other data from interfaces to external data providers. The data products comply with the FAIR principles and are computer-actionable, i.e. a computer system is able to find and access them and understand the corresponding schema. Data products are either physically available or built on demand when the product is accessed. In addition, self-service collection of community-agnostic tools are provided, such as Jupyter notebooks, which enable domain experts to create, deploy, and maintain community-specific applications on top of these data products.

The **Application Layer** consists of concrete applications and services developed for end users. These services can be community-agnostic, such as a search tool for datasets, or community-specific, such as a data portal for dragonflies or other species of interest. Community-specific applications are built on top of the data products in the semantic layer, while community-agnostic applications can access data from different layers. For example, the search tool requires access to data from the Cloud Layer and the Semantic Layer.

In addition to these four layers, there are two other essential elements in the architecture. The first one **Management & Governance** features tools and policies to manage rules and access rights for the resources offered in the four horizontal layers, including user management and monitoring of usage of the technical resources. The second, called **External Data Interfaces**, features a collection of interfaces for accessing external data sets. Obviously, RDC requires connectivity to established large data providers without the need to manage copies of their data in the Cloud Layer.

2.7.1 Linking Data and Service Providers with the NFDI RDC

Task Area 4 Measure 1 aims at the integration of data sources from data centres and data providers in multiple ways. First of all, the Task Area 4 has successfully integrated the GFBio Data Centers for both using their rich offerings of datasets and transferring datasets to them for the purpose of long archiving (Milestone 4.1.2). The GFBio Portal, GFBio Search, DMP Service and GFBio Submission Service offered as consortium services in the application layer of RDC are deployed on the de.NBI cloud and are interacting with the GFBio Data Centers. Secondly, RDC aims at integrating external datasets of institutions with high-quality and very large data repositories, e.g., IÖR-Monitor [50] and Earth Online [51], without having the data stored locally. Instead, RDC supports either FAIR interfaces of these repositories or, if not available, a thin and easy-to-implement protocol is going to be enabled for accessing external metadata [52] and data that need to be implemented by the

corresponding providers. So far, we have developed the design of such an interface, but implementations are not yet available.

2.7.2 Data Integration and Harmonization

This measure includes the development and integration of a new ontology service in the Mediation Layer of the RDC. BiodivPortal [32], [33] was developed by InfAI as a successor of the GFBio terminology service [45], using the international OntoPortal technology [53] and architecture as part of milestone 4.2.1. It is a repository that facilitates the management and accessibility of biodiversity related terminologies. It stores and manages a set of terminologies and standard schemas useful for transforming data and metadata into a standard format for processing and storage in the RDC Semantic Layer. BiodivPortal offers, among others, services to (i) annotate texts with terms and (ii) establish links between semantically related terms and (iii) make terms referenceable via stable URIs and (iv) make term versioning transparent.

2.7.3 Storage Infrastructure

The main achievements of this measure is the successful development of the Aruna Object Storage (AOS) [30], [31], [46] by the Justus-Liebig-University Gießen for Milestone 4.3.2. AOS is a native object storage that is used for storing datasets in a standardised manner obeying the FAIR data principles. Instead of organising datasets in a hierarchical file system, AOS manages them as objects with metadata and access rights attached. AOS supports accessing objects either via a RESTful API where JSON is used as the language for data exchange, or via the standardised S3 interface. AOS is intended as a data backend for the RDC. A workshop was held for milestone 4.3.1 to discuss the initial design of the RESTful API with software developers. Data import, verification, transformation and processing is basically possible via the services in the mediation layer. This also ensures the consistency of the data. Users and services can be informed about changes to individual data objects or even entire projects via the AOS notification service and can thus react to these changes.

2.7.4 Software Application Frameworks

Task Area 4 Measure 4 offers a first version of a Data Discovery Portal (DDP) [54] as an extension of the GFBio Search [55]. While the GFBio Search is used for discovery of datasets provided by the GFBio Data Centers, it needs to be adapted for searching the other datasets of RDC that are e.g. stored in AOS, data products created in the Semantic Layer or external datasets of other data providers. Thus, the goal of a general-purpose DDP is to make all datasets that are accessible in

the RDC findable. A first step of the DDP was to migrate the GFBio Search as a service in the underlying cloud. Furthermore, DDP was extended to make datasets findable beyond the ones of the GFBio repositories. A first implementation of DDP was tested with datasets from some of the use cases of Task Area 1 such as Gesellschaft deutschsprachiger Odonatologen e.V. (GdO), Gesellschaft für Ichthyologie e.V. (GfI), Leibniz-Institut für ökologische Raumentwicklung (IÖR) and Insekten Sachsen.

2.7.5 Governance and Monitoring

Within Task Area 4 Measure 5 (RDC Governance and Monitoring), a KPI dashboard called Scorpion has been implemented by the Leibniz Institute for Plant Genetics and Crop Plant Research (IPK) Gatersleben to enable KPI monitoring at a technical level (output for Milestone 4.5.4; Appendix D, Table D1). Scorpion (see Figure 6a) is designed as a central platform to collect and store the KPI set and its category assignments, the service KPI sets and the service KPI measurements. It is hosted by the IPK and implements a basic role-based access (RBAC) model, distinguishing between users and administrators. Administrators can manage the users, KPI measurements and generate report cards (see Figure 6c). Users can either visualise the KPI measurements per service or submit results to the dashboard via the graphical user interface (GUI) (see Figure 6b) or a RESTful API. For transparency, the results can be viewed by any user for any service.

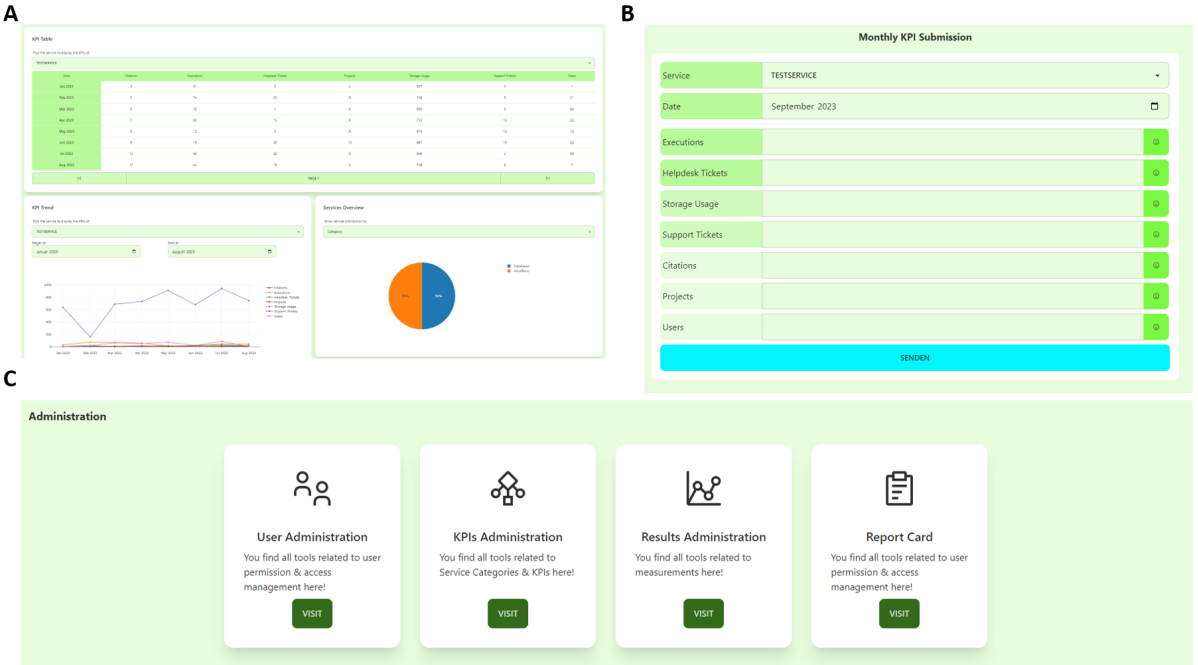

Figure 6: Overview for the Graphical User Interface of Scorpion

A) Home screen of the Scorpion Dashboard. The user gets access to KPI measurements in tabular form as well as a line chart. In addition, an overview either by category or by service provider is shown in the bottom right corner. B) For the monthly submission a GUI is provided, which can be used to submit KPI measurements monthwise to the system. The user selects the service and the month and the system displays the available KPIs. For each KPI an explanation text is provided by the tooltip on the right side. C) For administrators of the Scorpion Dashboard an administration view is provided, which allows the user to accept user's requests or remove erroneous KPI measurements.

In terms of a harmonised authentication (and potentially authorisation) infrastructure, we agreed on using LifeScience AAI [56], [57] as a joint solution (see Milestone 4.5.3; Appendix D Table D1). This solution is also being used in other Life Science Research infrastructures, like the German Bioinformatics Network de.NBI. We expect it to be compatible with the evolving IAM4NFDI, a basic service which is being developed across NFDI consortia. With the BIIGLE service (see BOX 3), a service piloted the implementation of LifeScience Login [26].

BOX 2 : Test pilot BIIGLE

The connectivity of services located in the Mediation Layer has been the subject of pilot testing. The service BIIGLE is a web-based software for image and video annotation that enables collaborative research on large datasets. Annotation is the process of marking objects of interest (e. g. with points, circles, polygons) in images or videos and describing these objects with labels. BIIGLE offers tools for manual and computer-assisted annotation, quality control and the collaboration on custom taxonomies to describe objects. The software is freely available for self-hosted installations or via the public instance which is maintained by the Biodata Mining Group at Bielefeld University. As a service for data analysis and visualisation, BIIGLE resides in the application layer of the RDC. To connect BIIGLE with the common infrastructure, Life Science Login is now supported to sign up for or log in to BIIGLE. Furthermore, users can now annotate images and videos stored in the Aruna Object Storage, which is developed as part of the RDC. Several software packages for the PHP language were developed as part of this process and can be reused in other projects.

3 Governance and Policies of the Service Portfolio

NFDI4Biodiversity is a practitioner network. Its members are committed to solutions which are suitable for everyday use and have been tried and tested in practice. In Task Area 3, we develop processes and measures that support service providers in offering sustainable, high-quality services and in making these services suitable for integration into the RDC. We also develop methods to monitor services. Many of these activities happen in close collaboration with other Task Areas and the SIG2 Service.

The SIG2 Service is a cross-task area group of experts tasked with designing and implementing a quality management strategy for NFDI4Biodiversity services. This includes deriving and establishing a set of Key Performance Indicators (KPIs) for usability and FAIRness, defining quality standards, and creating a service

documentation to track usage, user satisfaction, and resource requirements for each service. The certification process of the archives is also mentioned as a task area. The Service Coordinator, as chair of the SIG2 Service, is responsible for coordinating activities to implement the strategy in cooperation with the measures, as well as for the continuous quality assurance of the services. The position of Service Coordinator is part of the Project Coordination Office in Bremen.

A Service Quality Management strategy was developed to structure the assessment and selection of services for the NFDI4Biodiversity Service Catalog. Based on first experiences with integrating new services, a quality management concept is being developed and implemented. In this concept, services will be subjected to three phases of the Service Portfolio management: onboarding, monitoring and review. The processes will be managed in a Service Monitoring Committee, which includes the Service Coordinator as central manager and members of Task Area 3 - Long-term data, tools and service preservation (see Chapter 5).

3.1 Policy Development

For the provision of resources to the consortium, different roles were identified and defined (see Figure 7). The terms and definitions are oriented on EOSC [4]. Policies are being developed across the Task Areas to describe roles and responsibilities of the resource providers. They describe the general framework and the quality standards to be applied in the consortium for each type of resource provider. By separating policies and criteria, these documents can be created and maintained independently. The policies will be endorsed by the Steering Committee.

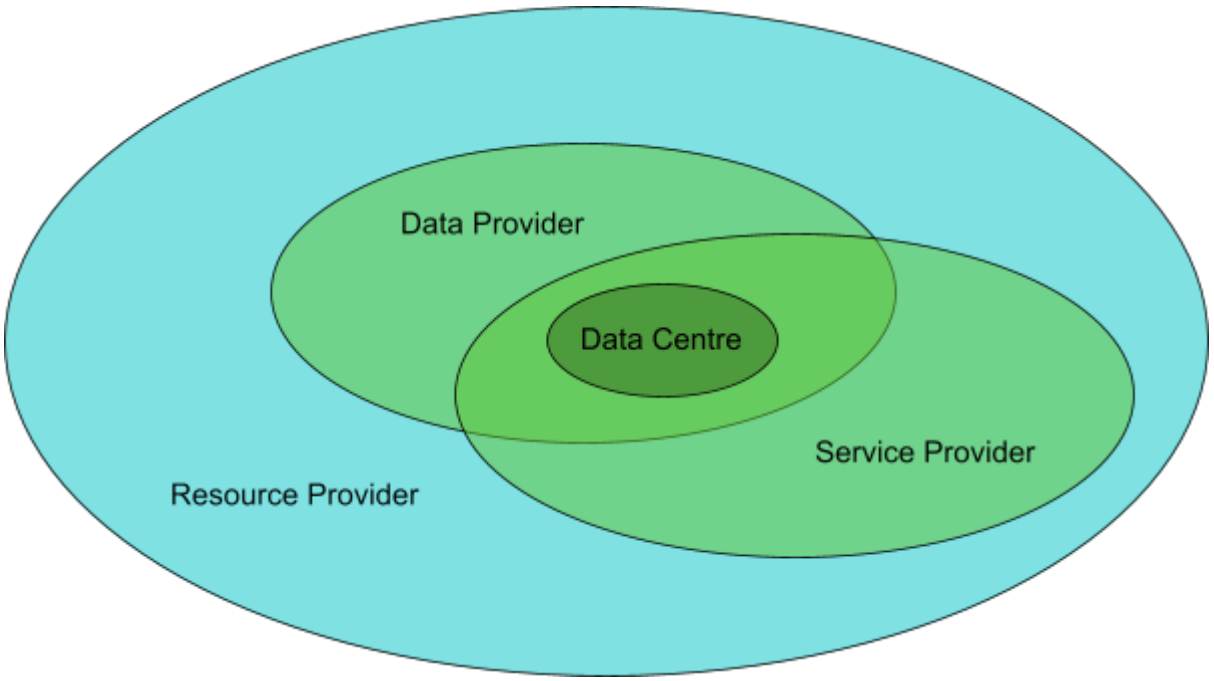

Figure 7: Resource Provider overview

Resource providers are participants offering digital objects or processes such as data and metadata, publications, software, workflows, services, and training materials as a resource for consumption. Data Providers are participants offering digital objects or processes such as data and metadata, publications, training materials as a resource for consumption. Service providers are participants offering digital objects or processes such as software, workflows, services as a resource for consumption. Data Centers are participants offering (external) services of a long-term digital archive/data repository as well as digital objects or processes as a resource for consumption.

In 2022 and 2023 a collection of blueprints and examples for provider policies was assembled and discussed in Resource Provider Workshops organised as part of Task Area 4 Measure 5 to define requirements for the policy development within NFDI4Biodiversity. Resulting from these workshops, following outputs have been defined:

- **Emphasise legal aspects:** It is important to clarify the legal framework under which the resource may be used. The consortium should encourage the use of open licences such as CC0 and CC-BY for data or MIT and GPLv3 for services. For new developments that are used in the Application Layer of the RDC open licences can be enforced. A dual- or multi-licensing model can be used if integration in the RDC is incompatible with the actual licence. For sensitive or embargoed data, the negotiation of data use agreements is possible.
- **Build on top of available solutions:** A solution for NFDI4Biodiversty must be compatible with preliminary works and should build on them. There are

common criteria for data providers [58] from the GFBio Project. In addition, the FAIR principles should be followed. For service providers, de.NBI has established guidelines [59] for service selection and quality management with KPI monitoring. In addition, most services are registered in registries such as bio.tools [60].

- **Guide the resource providers:** Resource providers need to be supported in their work so that they are willing to share their data or services. For this purpose, the added value of freely available data should be in the focus and explained. In order to foster a positive data sharing environment, the expertise of the data contributor must be recognized. On the service provider side, a well-defined, streamlined technology stack should be prioritised, as well as the use of platform-independent technologies wherever possible. This eases the maintenance and long-term deployment of the service.

Together with these requirements, Task Area 5 has compiled a collection of materials for further work on the Rules and regulations for RDC, which will serve as a basis for further elaboration.

A Service Provider Policy is being developed by Task Area 3 to enable uniform standards for high-quality services and to define and transparently communicate guidelines for the inclusion of services in the Service Portfolio and later in the Service Catalog. The development of the policy will be guided by existing solutions such as those of de.NBI, EOSC or DiSSCo.

A Data Provider Policy is being developed to guide data providers in achieving the goals of the NFDI4Biodiversity mission statement. The focus is on implementing the FAIR principles (see Chapter 4), developing high quality data offerings, and agreeing on guidelines for working together on and with data. A focus is set on the support that the consortium will provide to data providers in mobilising the data.

3.2 Onboarding of Services

Parallel to the initial assessment (see Chapter 2.2), an initial workflow for the onboarding process was developed by Task Area 3 and Task Area 1 Measure 2. To be included in the Service Portfolio, providers are requested to fill out a form, called BONSAI form (Biodiversity Onboarding Service Application Form). It contains all the information required for onboarding of a service including amongst others information about the service category, its provider, a description and available training and support. From this, a Service Profile will be (semi-)automatically generated, which provides a user-focused overview of the service. Information can also be extracted for governance and service monitoring purposes. In addition, providers are requested to provide Helpdesk support contacts and materials for outreach (PowerPoint slides and newsblog articles).

An onboarding process [61] was developed for two partner services, which served as pilot cases: BIIGLE (University of Bielefeld) and RightField (Heidelberg Institute for Theoretical Studies gGmbH). Currently, the onboarding of three services developed in the project (Aruna Object Storage, BiodivPortal (Terminology Service), Living Atlas Portal (LAND) [21]) is underway. The onboarding criteria and processes will be refined further to meet requirements related to the quality of project outcomes.

3.3 Continuous Service Monitoring

Following the onboarding process, a formal review process will be implemented to ensure that the performance of the services included in the NFDI4Biodiversity Service Catalog are regularly reviewed on their quality. This process will be based on the quality assurance approach described below (see Chapter 5). First proposals for a possible setup of this service have been developed by the Service Coordinator and will be discussed in SIG 2 and Task Area 3 before being presented to the Steering Committee for approval.

4 Implementation of the FAIR Principles

The Service Portfolio is instrumental in implementing the FAIR guiding principles for scientific data management and stewardship [7] , both at the community and the partner level. To this end, specific community-level developments have been supported, as well as several Use Case and flexfunds projects to advance and improve data mobilisation in partner organisations. At this stage, we define FAIR services as those that enable "working with and creating FAIR data". In a later phase, we want to systematically assess and improve FAIRness at the level of software and the services themselves.

4.1 Community Level

The NFDI4Biodiversity Community works with a portfolio of accepted metadata standards for data and digital objects in the field of taxonomy, natural history collections and genomics to enhance interoperability and findability [62]. Several consortium partners are engaged in standardisation activities, e.g. TDWG [63], the Genomics Standards Consortium/MIxS [64] or MIAPPE [65]. Partners mobilising occurrence data for GBIF and the Atlas of Living Germany [21] are encouraged to deliver the ABCD consensus elements agreed in the GFBio project [66]. Within the project, implementation of Bioschemas [67] was prioritised as the most promising standard to achieve convergence of the consortium in terms of discoverability of data resources, and increase potential for trans-disciplinarity (activity of Task Area 2 Measure 3). An exploratory workshop with data providers was conducted at the NFDI4Biodiversity All Hands Conference in 2022, followed by two hackathon-type

workshops in 2023, i.a. to improve findability of partner resources in search engines. These activities were supported by experts from DataPLANT and ELIXIR. The ambitious work plan for Task Area 2 Measure 3 was reduced after the initial budget cuts. Instead of the five milestones originally foreseen (including an extensive landscape analysis), a prioritisation exercise is conducted (starting from the compilation in [62], due in project month 29), and involvement of the consortium in relevant standardisation activities be brought together and documented (project month 29) and existing studies from recent years will be brought together and documented in a "meta analysis" (project month 42). The project consortium will not conduct its own standards ratification activities. If necessary, separate funding will have to be raised.

In terms of semantic integration, progress has been made to expand the GFBio Terminology service [45] into a centralised repository and service for semantic applications in NFDI4Biodiversity and beyond (Task Area 4 Measure 2). Caching mechanisms to centrally store essential decentralised resources have been implemented as ETL (Extract Transform Load) processes. In particular, resources useful for the portal semantic search (work in progress: M4.2.4 adaptable semantic integration) are being prioritised.

We carry on linking our efforts with European and international initiatives and infrastructures. The partner InfAI is engaged in the Ontoportal Alliance [53] and has replaced the backbone of the GFBio Terminology Service by an OntoPortal installation (see Chapter 2.7.2, [32], [33]). Beside contributing to national (see Box 3) and international collaborative efforts, we are involved at an early stage in standardisation efforts for terminology evolution [68] as well as the adoption of standards for terminology mapping [69]. This will reinforce our work on terminology mapping and linking capabilities [70], [71] and versioning and synchronisation mechanisms [72].

BOX 3 : Terminology Service for NFDI

FAIR research data requires formal and common terminologies, for the purposes of knowledge representation, enhancing data discovery, consistency, and integration in the NFDI context. NFDI consortia work with terminologies at different levels of maturity and FAIRness, their alignment, management, and development are challenging tasks. Terminology services (TSs) have been developed to address these issues, they however still face challenges with interoperability and widespread adoption. The Terminology Services 4 NFDI (TS4NFDI) vision aims to standardise and harmonise terminology management within the NFDI, enabling cross-disciplinary interoperability and encouraging terminological alignment for standardised, interoperable, and sustainable services and knowledge engineering frameworks.

Partners involved in Task Area 4 Measure 2 - Data integration and harmonisation continue developing semantic metadata exchange formats based on internationally agreed semantic standards that are required by the RDC Semantic Storage (Task Area 4 Measure 3) for data linkage and schema transformation. The initial catalogue of metadata models has been expanded with semantic core models

including Schema.org, ABCD 3.0, Dublin Core and Datacite. The adoption and extension of such models is done in close cooperation with the community. In particular, the need for a more fine grained description of observable properties led to the adoption of the RDA I-Adopt [73] (InteroperAble Descriptions of Observable Property Terminology) semantic framework and its integration to the metadata model of Schema.org. A slice of I-Adopt terminologies is made available via BiodivPortal and an automatic annotation is provided through the service to enable an automatic generation of I-Adopt metadata annotations [74].

4.2 Partner Level

The schema for mobilisation of species occurrence data from Use Case partners highlights how best practices already available in the community are being applied to implement the FAIR principles:

- The **findability** of data is being improved by publishing datasets in a suitable GFBio Data Center with supportive curation staff. Data is then pushed to the Global Biodiversity Information Facility (GBIF) and to the GFBio portal, with their respective community-oriented search indexes.
- The deposit in data archives also improves **accessibility**: The data archives issue persistent identifiers for each data set and ensure that the data producers specify the access rights by a clear licence.
- To increase **interoperability** of the species occurrence data, providers are supported in delivering the ABCD consensus elements with the metadata [66] and to use standardised ontologies developed as part of the Research Data Commons in Task Area 4 [33]. The IT infrastructure of the GFBio Data Centers provide an API to ensure machine-actionability of the metadata.
- To increase **reusability**, data providers are encouraged to use the Creative Commons licences CC0 and CC-BY, in cases where the German Urheberrecht (copyright) laws apply.
- In this way, implementation of FAIR data practices has been supported for the Red list of freshwater fish and data from the fish atlas of the German Ichthyological Society (GfI e.V.) [75]

Several partners were able to improve FAIR data practices through dedicated Use Case and flexfunds projects

- The Arachnological Society (AraGes) cooperates with the Bavarian Natural History Collections (SNSB) and the State Museum for Natural History in Karlsruhe (SMNK) to manage data in a state-of-the art data management system (Diversity Workbench) [76]. A flexfunds project granted in 2022 improved the findability and accessibility of individual datasets within the ARAMOB database (e.g. all observations from a common project or sampling

campaign). To do this, AraGes cooperated with the SNSB to extend the concepts for DiversityProjects and other modules of the Diversity Workbench (DWB) system which resulted in new DWB implementations. They enriched the DWB metadata with project-level information. Specific search filters were set up for the ARAMOB portal. This work lays the foundation for planned data publications through the journal "Arachnologische Mitteilungen". By encouraging authors to submit data publications, AraGes and ARAMOB will realise a cooperation with the GFBio Data Centers Stuttgart State Museum of Natural History (SMNS) and SNSB to mobilise more datasets in future for GBIF and NFDI via SNSB BioCASe provider software API.

- The Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) conducted a pilot project to mobilise biodiversity sequencing data according to the ELIXIR-MIAPPE Use Case [77]. As a result, 2,500 records have been registered at EMBL-BioSamples. Furthermore, 12 IPK biodiversity projects have been registered at EBML-ENA. In total an amount of 39 Terabyte sequence raw read files could be transferred.
- Friedrich-Schiller-Universität Jena used a flexfunds project to further improve the FAIRness of the BEXIS2 workbench. In particular, with an extended metadata API, interoperability could be improved. A further significant achievement is the new export of data as the Darwin Core Archive. With this, it becomes possible, with low effort, to make data held in BEXIS2 available in GBIF and thus increase its finability for the community.
- The Federation of German Avifaunists (DDA) holds large and heterogeneous datasets on the occurrence and distribution of birds in Germany. The flexfunds enabled a new use case project dedicated to making these datasets more accessible. DDA's dbird [78] database is being mapped onto the ABCD metadata standard, in order to improve interoperability and reusability. A research data portal is being developed to facilitate requests for research data sets from ornitho.de.
- In cooperation with the Leibniz Institute of Ecological Urban and Regional Development (IÖR), the Archives of the Bavarian State were able to annotate historical maps with additional metadata. For this purpose, the historical maps are mobilised and digitised within the Archives of the Bavarian State and integrated into their own archiving system with UUID. With the planned ABCD interface in the archiving software (currently in the test phase), the ABCD annotations extracted from the IOER software can be integrated.
- The Johann Heinrich von Thünen Institute extended the GeoNode [27], [79] software by improving its API, which covers several use cases [80]. GeoNode is developed as an open-source FAIR data portal for spatial datasets and the project financed by the NFDI4Biodiversity consortium laid the foundation for further development of this infrastructure component. The outputs were communicated in workshops [81] and conferences [82].

4.3 Cross-links to other Projects

One of the GFBio Data Centers, the data publisher PANGAEA, is a cooperation partner in the FAIRsFAIR project on the European level and has been confirmed to have a high degree of FAIRness according to the Fuji-Tool [83]. The other GFBio Data Centers are not conducting FAIR assessment exercises at this point.

NFDI4Biodiversity is a use case in the BMBF-funded project “FAIR Data Spaces” (FAIR-DS), which is conducted in cooperation with the NFDI directorate [47]. The project is designed to create a roadmap for the collaboration of NFDI GAIA-X (2021-2024). We expect synergies with Task Area 4 Measure 5 in the clarification of the ethical and legal framework for data exchange between research and industry, and with Task Area 4 in general, regarding the technical foundation and demonstration of Gaia-X technology for providing and using research data according to the FAIR Principles.

During the creation of the Service Profiles, additional references were examined; there were overlaps between information already included in the Service Profile and the RDA reference criteria for FAIR Research Software [84]; these can later be elaborated as an additional measure of FAIRness.

5 Development of Quality Assurance

To offer a Service Catalog composed of trustworthy and high-quality services, it is essential to establish processes for ensuring and managing quality. These processes should be guided by Key Performance Indicators (KPIs) to serve as critical benchmarks.

5.1 Development of a Quality Management Strategy

SIG2 Service has been tasked by the steering committee to develop a comprehensive yet lean quality strategy for the services provided by the consortium partners. In March, SIG2 Service held two workshops to outline the key points for the quality strategy.

The concern regarding structured quality assurance straining limited resources requires a stepwise implementation of quality management. Services funded by projects and partner-operated ones both hold responsibility for ensuring quality,

overseen by the steering committee's mandate for comprehensive quality assurance. The diverse spectrum of the service providers in the consortium, ranging from IT centres to volunteer organisations, necessitates flexible, concept-based quality management due to varying prerequisites. While the consortium can foster governance, the ultimate responsibility for service quality lies with the provider. To ensure compliance with requirements, a detailed description of roles and processes within the project is essential. The helpdesk acts as a central point for users, requiring the allocation of resources for involvement by service providers and training for uniform responses. Transparency, emphasised in the NFDI4Biodiversity Mission Statement, fosters trust internally and externally.

Based on these workshop results, a draft service quality management strategy was presented to the Steering Committee in April 2023:

The primary quality goal: The quality management focuses on an overall goal for a strategic period, from which the objectives are derived. The primary goal for the current phase of the project, derived from the mission statement, is "To establish the consortium as a reliable infrastructure provider, recognized for the competence of its contacts and the transparency of its services."

Developmental goals for quality management: In order to take into account the resources and the level of development of the consortium, the development of the quality management must be done successively. Therefore the consortium needs development goals for its quality management for each strategy periode.

Quality objectives: To ensure a consistent level of quality and reputation across the consortium and thus build trust. Three concrete quality objectives were derived from the primary quality goal:

1. Focus on services for the personas [85] "Paul" (Post-Doc) and "Doro" (data manager).

2. External branding as a consortium with strong user support (through the Helpdesk).

3. Internal branding through practical support of partner institutions in quality management (templates, etc.).

Service Areas: The consortium is centred on a wide range of services, including software-based specialised applications and offerings such as training and

consultation, each characterised by unique user expectations for their quality attributes. To address these demands effectively, the proposal is to organise quality management activities into three distinct service domains so-called service areas:

- **Service Area: Tools** - includes databases, web applications, APIs, and other specialised applications that support data provisioning, management and processing, analysis, as well as the components of the RDC architecture beneath the Application Layer. It entails specific quality requirements for technically oriented services.
- **Service Area: Training and Education** - encompasses the consortium's portfolio of teaching, further education, and training offerings, with specific requirements for the quality of teaching and learning.
- **Service Area: Helpdesk** - comprises the consortium's collectively organised support services for research data management, certification, and tools. It encompasses specific quality requirements for advisory servicesP

Quality assurance and controlling: For each of the service areas a quality assurance concept will be developed, including three process for quality assurance and controlling:

1. Onboarding and Evaluation: Services go through an onboarding process involving an evaluation of their quality, confirming compliance with essential features.
2. Ongoing Monitoring: Service providers engage in continuous monitoring to ensure quality assurance.
3. Performance Review: Services undergo periodic performance reviews, potentially annually, to enhance and refine their quality.

Proposed responsibilities include a Quality Officer (a position to be filled by the Service Coordinator), leads for each service area, and distinct roles for evaluation, monitoring, and performance reviews.

The quality strategy will be presented in the Task Areas and Measures and the overarching SIG5 Strategy. The development of quality assurance concepts for each service area will follow, describing quality planning, evaluation, and improvement, and expanding the onboarding process to include service quality evaluation. The integration of KPI from initial service providers into the Scorpion dashboard will be completed by September 2023. Long-term goals involve concepts and resources for usability testing and continuity planning.

5.1 Guidance on Certification

For data providers who aim to get a certification for their repository, the consortium offers support in form of a Data Center peer group, collected information in the project plattform [86], having generic templates in place and a specialised Helpdesk support.

NFDI4Biodiversity supports a CoreTrustSeal certification for repositories. Any CTS certified provider in the field can become a data centre, if they meet the domain-specific criteria under development of the German Federation for Biological Data (GFBio e.V.). In time, support with certifications may be extended for other types of services, to support interested service providers accordingly.

5.2 Monitoring of Service Quality

The service quality monitoring is planned for all services of the Service Catalog. There are two types of measures for the quality assurance of services: Besides the qualitative measures that are planned, a quantitative KPI measuring is implemented [87]. We intend to start monitoring with pilot service providers this year and review the concept after the first results are available.

Practically, the Service Monitoring Committee and the KPI monitoring (Task Area 4 Measure 5) jointly select a service category together with the service provider that defines its primary KPI set. For the definition of the initial KPI set, the preliminary work of de.NBI [9] was used and the KPIs were unified. In addition, the necessity of the KPIs was adapted to the use in NFDI4Biodiversity. With the KPIs defined, sets for each category were formed (see Figure 8). In addition, service providers can decide to add optional KPIs exceeding the KPIs defined by their service category.

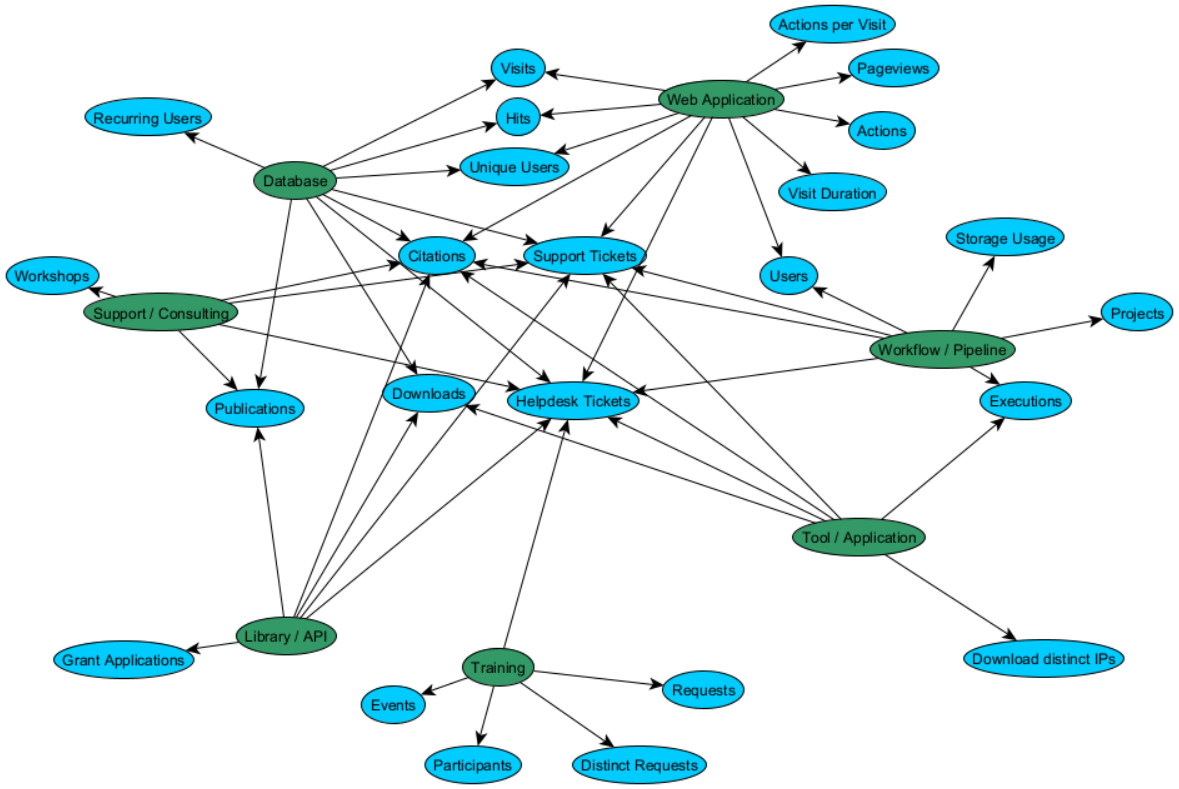

Figure 8: Graph representation of the KPIs and the predefined KPI sets of the categories.

Each category (green node) is linked with an associated necessity to at least one KPI (blue node). A service's KPI set then comprises the KPIs as defined by their category as well as all optionally added KPIs.

5.3 Sustainability Assessment

The NFDI4Biodiversity consortium is well aware that an important element for the acceptance of its service portfolio is sustainability. Sustainability at present means the continuous maintenance and operation of technical and organisational structures and services, including adaptations to changing scientific and technical needs. NFDI4Biodiversity's service-oriented RDC architecture, building on modular, re-usable, and exchangeable components, is designed to increase the sustainability of the joint infrastructure and service components. In accordance with the cooperation contract, components developed within the project will be open source and in compliance with international standards. All project related materials and results will be made publicly available.

Currently, few of the partners have mandates to run services for the national/international community, but rather open their tried-and-tested infrastructure and services for external users on an informal basis. In particular this often means that core services (like the collection data centres and PANGAEA)

are stable, but the staff for community services (data curators, developers etc.) is mainly project-funded. Additionally, services to be integrated are often heterogeneous with respect to their legal status, technical and organisational structures, and funding. This not only increases the effort required for the operation of common services but also makes it more difficult setting up a common legal framework and leads to an unbalanced commitment of members.

To provide a long-term service for providers and users who want to archive, publish and discover data, Task Area 5 Measure 2 investigates the NFDI4Biodiversity Service Catalog in terms of potential costs and sustainability models, taking into account the working conditions of the institutions involved. To get a handle on the current sustainability status of the Service Portfolio of NFDI4Biodiversity a set of basic questions to be answered by the partners and service providers have been extracted from the requirements catalogue of the CoreTrustSeal certification process [16]. The questions are slightly different for services/software and repositories/data archive (see Appendix C).

The questionnaire was discussed in the Special Interest Group 5 - SIG Strategy and information requests included in the Service Profile Template (see Appendix B). An analysis will be carried out when the Service Portfolio documentation is completed.

It should be noted that we have no intentions to monetise the services we provide. Based on the experiences and model calculations we did in the GFBio project we would even state that “selling” data or services does not lead to significant revenues in an academic environment.

In summary NFDI4Biodiversity has started to address and better understand the sustainability issues of its service providers. Besides managing the needs of the partners in terms of funding the managing partner GFBio e.V. has started to earmark and take over the responsibility of a couple of core services for the provision of long-term services.

6 Outlook

Three years into the initial funding phase of NFDI4Biodiversity, we have significantly progressed in harmonising data-related services across providers in Germany. Future tasks of the NFDI4Biodiversity Consortium can be divided into two key areas: Development of a coherent Service Catalog for stakeholders in the biodiversity domain and development of pilot applications for the Research Data Commons.

Regarding the Service Catalog, the aim is to provide selected, tried-and-tested services from the portfolio with additional support and offer them to the Community under a joint quality management framework. The support enables service providers to implement the quality assurance framework and to implement

common standards (e.g. in terms of software / product quality). Regular reviews will serve to improve the established processes relevant to achieve service quality. For the KPI monitoring we will explore possibilities for Scorpion to harvest the KPI measurements instead of a push approach as well as the automation of creation of summary statistics of the Service Monitoring for the review in Scorpion and the development of an interactive, wizard-like solution for (semi-)automated service onboarding, based on the BONSAI form. Many of our services enable community stakeholders to work with and create FAIR data. In a later phase, we want to systematically assess and improve FAIRness at the level of software and the services themselves.

In the area of RDC development, the documentation currently being updated will be at a scale that will allow for further coordination of service development as well as connecting further pilot applications to the layers of the architecture. Based on current collaboration, we can confirm that a self-service platform as RDC would meet specific user needs, especially as a combination of own applications and compute/storage capacities. Ideally, a multi-cloud environment and professional backup infrastructure would be provided. A productive RDC platform however would ultimately need to be rolled out and supported across NFDI consortia. We will therefore explore cooperation with other consortia and perspectives within the Base4NFDI initiative.

The quality management strategy developed in 2023 will enable us to survey user satisfaction in a structured way, which is relevant for both areas of development. Regular quality surveys and user feedback will be set up, and collected user needs will be translated into features and evaluated in terms of user satisfaction and service performance.

References

[1] „Start page - NFDI4Biodiversity“. https://www.nfdi4biodiversity.org/en/ (assessed: 15.08.2023).

[2] „nfdi | Nationale Forschungsdateninfrastruktur e. V.“ https://www.nfdi.de/ (assessed: 5.09.2023).

[3] K. des V. N. F. (NFDI) e.V, „Stellungnahme der NFDI-Konsortien zu Basisdiensten“, Feb. 2022, doi: 10.5281/zenodo.6091657.

[4] Directorate-General for Research and Innovation (European Commission) und EOSC Executive Board, EOSC rules of participation. LU: Publications Office of the European Union, 2021. Assessed: 15.08.2023. [Online]. Available at: https://data.europa.eu/doi/10.2777/30541

[5] „RDMkit“. https://rdmkit.elixir-europe.org/data_life_cycle (assessed: 31.08. 2023).

[6] „Code of Conduct - Wissenschaftliche Integrität“. https://wissenschaftliche-integritaet.de/en/code-of-conduct/ (assessed: 15.08.2023).

[7] M. D. Wilkinson u. a., „The FAIR Guiding Principles for scientific data management and stewardship“, Sci. Data, Bd. 3, Nr. 1, Art. Nr. 1, March 2016, doi: 10.1038/sdata.2016.18.

[8] „Mission Statement - NFDI4Biodiversity“. https://nfdi4biodiversity.org/en/mission-statement/ (assessed: 15.08.2023).

[9] M. Turewicz, U. Scholz, F. O. Glöckner, T. Dammann-Kalinowski, und M. Wittchen, de.NBI service category-specific KPI selection and criteria. Zenodo, 2022. doi: 10.5281/zenodo.6597826.

[10] L. Amelung u. a., „White Paper: Interim Report Reference“, Zenodo, March 2023. doi: 10.5281/zenodo.7688729.

[11] „BIIGLE | Next generation image and video annotation“. https://biigle.de/ (assessed: 15.08.2023).

[12] „RightField | Semantic data annotation by Stealth“. https://rightfield.org.uk/ (assessed: 15.08.2023).

[13] „Data Publisher for Earth & Environmental Science“. https://www.pangaea.de/ (assessed: 15.08.2023).

[14] „BfN Checklist API“. https://checklisten.doctronic.de/api/public/swagger-ui (assessed: 15.08.2023).

[15] „Contact - NFDI4Biodiversity“. https://www.nfdi4biodiversity.org/en/kontakt/ (assessed: 15.08.2023).

[16] C. S. and C. Board, „CoreTrustSeal Trustworthy Data Repositories Requirements: Glossary 2020–2022“, Nov. 2019, doi: 10.5281/zenodo.3632563.

[17] „GFBio | German Federation for Biological Data - Services“. https://www.gfbio.org/services/ (assessed: 15.08.2023).

[18] „What we do: Use cases overview- NFDI4Biodiversity“. https://www.nfdi4biodiversity.org/en/what-we-do/use-cases-uebersicht/ (assessed: 15.08.2023).

[19] „Basic Ideas behind ITIL | IT Process Wiki“, IT Process Wiki - the ITIL® Wiki. https://wiki.en.it-processmaps.com/index.php/Basic_Ideas_behind_ITIL (assessed: 15.08.2023).

[20] B. Ebert, J. Engel, I. Kostadinov, A. Güntsch, und F. O. Glöckner, „Connecting National and International Data Infrastructures in Biodiversity Research: The Case of NFDI4Biodiversity, a German Consortium for Biodiversity, Ecology and Environmental Data“, gehalten auf der CoRDI Conference 2023, Karlsruhe. doi: https://doi.org/10.52825/CoRDI.v1i.346.

[21] „Lebendiger Atlas - Natur Deutschland (LAND) - Lebendiger Atlas - Natur Deutschland“. https://land.hp.gbif-staging.org/ (assessed: 15.08.2023).

[22] P. Koss Piepenburg, Dieter, Teschke, Katharina, Dannheim, Jennifer, Brey, Thomas, „Current State of Use Case CRITTERBASE“, 30. May 2023. doi: 10.5281/zenodo.7984765.

[23] „VAT: Visualization, Analysis and Transformation System“. https://vat.gfbio.org/#/ (assessed: 16.08.2023).

[24] M. Zurowietz, „biigle/flysystem-aruna: v1.0“. Zenodo, 17. May 2023. doi: 10.5281/zenodo.8059139.

[25] M. Zurowietz, „biigle/laravel-aruna-driver: v1.1“. Zenodo, 17. May 2023. doi: 10.5281/zenodo.8059145.

[26] M. Zurowietz, „biigle/auth-lslogin: v1.0.0“. Zenodo, May 2nd 2023. doi: 10.5281/zenodo.8059127.

[27] „Informationen für Entwickler — Thünen Atlas“. https://atlas.thuenen.de/developer/ (assessed: August 2023).

[28] D. S. Domisch, „The hydrographr R-package“, Global freshwater biodiversity research, 5. June 2023. https://glowabio.org/project/hydrographr/ (assessed: 15.08.2023).

[29] „hydrographr“. Global freshwater biodiversity research, 2. August 2023. Assessed: 15.08.2023. [Online]. Available at: https://github.com/glowabio/hydrographr

[30] „ArunaStorage“, GitHub. https://github.com/ArunaStorage (assessed: 15.08.2023).

[31] „Aruna Object Storage“. https://aruna-storage.org/ (assessed: 15.08.2023).

[32] „BiodivPortal“, GitHub. https://github.com/biodivportal (assessed: 15.08. 2023).

[33] „RDC Mediation Layer: BiodivPortal - NFDI4Biodiversity - GFBio Confluence“. https://kb.gfbio.org/display/NFDI/RDC+Mediation+Layer%3A+BiodivPortal (assessed: 15.08.2023).

[34] „Service Monitoring: Scorpion - NFDI4Biodiversity - GFBio Confluence“. https://kb.gfbio.org/display/NFDI/Service+Monitoring%3A+Scorpion (assessed: 28.08.2023).

[35] „Software – Diversity Workbench“. https://diversityworkbench.net/Portal/Software (assessed: 5.09.2023).

[36] „BEXIS2“. https://bexis2.github.io/ (assessed: 15.08.2023).

[37] T. Astor, J. Weber, I. Kostadinov, F. O. Glöckner, und J. Nieschulze, „Potential für ein starkes Netzwerk zwischen GFBio und FDM-Beratenden an Universitäten und Forschungsinstituten“, Bausteine Forschungsdatenmanagement, Nr. 1, Art. Nr. 1, March 2021, doi: 10.17192/bfdm.2021.1.8311.

[38] „Home · FEdA – Die BMBF-Forschungsinitiative zum Erhalt der Artenvielfalt“. https://www.feda.bio/de/ (assessed: 15.08.2023).

[39] „Willkommen - Qualiservice Data Sharing“. https://www.qualiservice.org/de/ (assessed: 15.08.2023).

[40] „Data Centers - NFDI4Biodiversity - GFBio Confluence“. https://kb.gfbio.org/display/NFDI/Data+Centers (assessed: 15.08.2023).

[41] „GFBio | German Federation for Biological Data - Data Centers“. https://www.gfbio.org/data-centers/

[42]„GFBio | German Federation for Biological Data“. https://www.gfbio.org/

[43] C. Hassenrück, T. Poprick, V. Helfer, M. Molari, R. Meyer, und I. Kostadinov, „FAIR enough? A perspective on the status of nucleotide sequence data and metadata on public archives“. bioRxiv, S. 2021.09.23.461561, 24. September 2021. doi: 10.1101/2021.09.23.461561.

[44] „CoreTrustSeal“, CoreTrustSeal. https://www.coretrustseal.org/ (assessed: 15.08.2023).

[45] „GFBio | German Federation for Biological Data - Terminology Service“. https://www.gfbio.org/terminologyservice/ (assessed: 15.08.2023).

[46] M. A. Dieckmann, S. Beyvers, A. Rehm, F. Förster, und A. Goesmann, „The Aruna Object Storage - A distributed multi cloud object storage system for scientific data management“. Assessed: 15.08.2023. [Online]. Available at: https://drive.google.com/file/d/1FAwUkypRjLMCEyhdY7gLQn2DDsmqcq7o/view?usp=sharing&usp=embed_facebook

[47] „FAIR Data Spaces | NFDI“. https://www.nfdi.de/fair-data-spaces/?lang=en (assessed: 15.08.2023).

[48] M. Diepenbroek, S. Schimmler, und B. Ebert, „Sektionskonzept Common Infrastructures zur Einrichtung einer Sektion im Verein Nationale Forschungsdateninfrastruktur (NFDI) e.V.“, Oct. 2021, doi: 10.5281/zenodo.5607490.

[49] „Data Mesh Principles and Logical Architecture“, martinfowler.com. https://martinfowler.com/articles/data-mesh-principles.html (assessed: 30.08. 2023).

[50] „Monitor der Siedlungs und Freiraumentwicklung“. https://monitor.ioer.de/?baselayer=topplus&opacity=0.8&raeumliche_gliederung=gebiete&zoom=6&lat=51.32374658474385&lng=10.458984375000002&time=2022&basiskarte=true&glaettung=0& (assessed: 28.08.2023).

[51] „Data - Earth Online“. https://earth.esa.int/eogateway/catalog (assessed: 28.08.2023).

[52] C. Beilschmidt, D. Brandenstein, J. Drönner, N. Glombiewski, M. Mattig, und B. Seeger, „On the Design and Implementation of Easy Access to External Spatiotemporal Datasets in NFDI“, held at the CoRDI Conference 2023, Karlsruhe.

[53] O. A. (Clement J. Graybeal) John, „OntoPortal“, OntoPortal. https://ontoportal.org/ (assessed: 15.08.2023).

[54] M. Feser, C.-T. Pfaff, M. Weber, I. Kostadinov, und U. Scholz, „TA4M4: Frameworks für Software Applikationen“, 12. October 2022. doi: 10.5281/zenodo.7589434.

[55] „GFBio Data Search“. https://search.gfbio.org/ (assessed: 15.08.2023).

[56] „Authentication Services: Life Science Login - NFDI4Biodiversity - GFBio Confluence“. https://kb.gfbio.org/display/NFDI/Authentication+Services%3A+Life+Science+Login (assessed: 30.08.2023).

[57] „Home | LifeScience RI“. https://lifescience-ri.eu/home.html (assessed: 30.08.2023).

[58] „Eligibility criteria for GFBio portal data providers – GFBio Public Wiki“. https://gfbio.biowikifarm.net/wiki/Eligibility_criteria_for_GFBio_portal_data_providers (assessed: 15.08.2023).

[59] de.NBI, „de.NBI/ELIXIR-DE Guidelines for Node Services Selection Process and Quality Management“, Feb. 2020, doi: 10.5281/zenodo.3967215.

[60] „bio.tools · Bioinformatics Tools and Services Discovery Portal“. https://bio.tools/ (assessed: 28.08.2023).

[61] „Entwurf Service Aufnahmekriterien - NFDI4Biodiversity - GFBio Confluence“. https://kb.gfbio.org/display/NFDI/Entwurf+Service+Aufnahmekriterien (assessed: 15.08.2023).

[62] „Data exchange standards, protocols and formats relevant for the collection data domain within the GFBio network – GFBio Public Wiki“. https://gfbio.biowikifarm.net/wiki/Data_exchange_standards,_protocols_and_formats_relevant_for_the_collection_data_domain_within_the_GFBio_network

[63] „Biodiversity Information Standards (TDWG)“. https://www.tdwg.org/ (assessed: 15.08.2023).

[64] „Background - MIxS“. https://genomicsstandardsconsortium.github.io/mixs/introduction/background/ (assessed: 15.08.2023).

[65] „Minimal Information About Plant Phenotyping Experiment“. https://www.miappe.org/

[66] „ABCD for data publication - GFBio Public Wiki“. https://gfbio.biowikifarm.net/wiki/ABCD_for_data_publication

[67] „Bioschemas - Bioschemas“. http://bioschemas.org/ (assessed: 15.08. 2023).

[68] „KGCL“. https://incatools.github.io/kgcl/ (assessed: 15.08.2023).

[69] Matentzoglu, Nicolas, James P. Balhoff, Susan M. Bello, Chris Bizon, Matthew Brush, Tiffany J. Callahan, Christopher G. Chute, u. a. „A Simple Standard for Sharing Ontological Mappings (SSSOM)“. arXiv, 13. Dezember 2021. https://doi.org/10.48550/arXiv.2112.07051.

[70] Alsayed Algergawy, Naouel Karam, Amir Laadhar, und Franck Michel, „Too big to match: a strategy around matching tasks for large taxonomies. Proceedings of the 17th International Workshop on Ontology Matching (OM 2022) co-located with the 21th International Semantic Web Conference (ISWC 2022)“, in CEUR Workshop Proceedings 3324, Hangzhou, China, Oct. 2022. [Online]. Available at: CEUR-WS.org

[71] Mina Abd Nikooie Pour u. a., „Results of the Ontology Alignment Evaluation Initiative 2022. Proceedings of the 17th International Workshop on Ontology Matching (OM 2022) co-located with the 21th International Semantic Web Conference (ISWC 2022)“, in CEUR Workshop Proceedings 3324, Hangzhou, China, Oct. 2022. [Online]. Available at: CEUR-WS.org

[72] Andreas Kohlbecker, Naouel Karam, Adrian Paschke, und Anton Güntsch, „Preserving Taxonomic Change and Subsequent Taxon Relationships over Time. Proceedings of the Joint Ontology Workshops (JOWO) 2021 Episode VII: The Bolzano Summer of Knowledge“, in CEUR Workshop Proceedings 2969, Bolzano, Italy, Sep. 2021. [Online]. Available at: CEUR-WS.org

[73] Magagna, Barbara, Ilaria Rosati, Maria Stoica, Sirko Schindler, Gwenaelle Moncoiffe, Anusuriya Devaraju, Johannes Peterseil, and Robert Huber. „The I-ADOPT Interoperability Framework for FAIRer data descriptions of biodiversity“. arXiv, 14. Juli 2021. https://doi.org/10.48550/arXiv.2107.06547.

[74] Karam, Naouel. „From the GFBio Terminology Service to BiodivPortal: A Terminology Repository and Service for Biodiversity with the OntoPortal Technology“, Bd. https://www.tib-op.org/ojs/index.php/CoRDI/index, 2023.

[75] M. Friedrichs-Manthey und J. Freyhof, „Harmonised freshwater fish occurrence and abundance data for 12 federal states in Germany“, doi: 10.15468/c75fky.

[76] „Publication of Type 1 Data via BioCASe Data Pipelines at SNSB Data Center – GFBio Public Wiki“. http://gfbio.biowikifarm.net/wiki/Publication_of_Type_1_Data_via_BioCASe_Data_Pipelines_at_SNSB_Data_Center (assessed: 15.08.2023).

[77] „Plant genomic and genetic variation data submission to EMBL-EBI databases“. https://fairplus.github.io/the-fair-cookbook/content/recipes/reusability/miappe.html (assessed: 15.08.2023).

[78] „dBird“. https://dbird.org/ (assessed: 15.08.2023).

[79] „GeoNode“. GeoNode Development Team, 27. August 2023. Assessed: 28.08.2023. [Online]. Available at: https://github.com/GeoNode/geonode

[80] „API usage examples — GeoNode master documentation“. https://docs.geonode.org/en/master/devel/api/usage/index.html (assessed: 5.09.2023).

[81] „GeoNode - Anwendertreffen FOSSGIS-Konferenz 2023“. https://pretalx.com/fossgis2023/talk/WYXSSC/ (assessed: 5.09.2023).

[82] „Neues aus dem GeoNode Projekt - TIB AV-Portal“. https://av.tib.eu/media/61194 (assessed: 5.09.2023).

[83] DANS, DCC, Directorate-General for Research and Innovation (European Commission), EFIS, und Visionary Analytics, „3.2.4 Barriers, motivators and enablers“, in European Research Data Landscape: final report, LU: Publications Office of the European Union, 2022, S. 36. Assessed: 15.08.2023. [Online]. Available at: https://data.europa.eu/doi/10.2777/3648

[84] N. P. Chue Hong u. a., „FAIR Principles for Research Software (FAIR4RS Principles)“, Mai 2022, doi: 10.15497/RDA00068.

[85] „Personas - NFDI4Biodiversity“. https://www.nfdi4biodiversity.org/en/personas/ (assessed: 15.08.2023).

[86] „CoreTrustSeal Certification - NFDI4Biodiversity - GFBio Confluence“. https://kb.gfbio.org/display/NFDI/CoreTrustSeal+Certification (assessed: 15.08.2023).

[87] „General · IPK-BIT/scorpion“, GitHub. https://github.com/IPK-BIT/scorpion (assessed: 5.09.2023).

Appendix

A. Services provided by Consortium Partners

Table A1 - SERVICE PORTFOLIO AS OF AUGUST 2023

Overview of the services provided by the consortium partners with a short description of the partner's expertise. Printed in bold are the services that indicated a Helpdesk contact.

Service Provider	Description	Services
Alfred Wegener Institute – Helmholtz Centre for Polar and Marine Research	AWI is a strong partner for the exchange of data and data products with NFDI4Earth. It will contribute to international standardization committees as well as ontology development. It will link out to EBV, NEON and ESIP. It contributes to use case 5.	**O2A**; CRITTERBASE
Arachnological Society	AraGes-Spinnen is a natural history society (spiders) leading use case 16 in a formal cooperation with SMNK - Staatliches Museum für Naturkunde Karlsruhe). Use case 16 connects the society's resources, including monitoring and collection data about spiders, with NFDI4BioDiversity. SMNK is contributing data curatorial and technical services and expertise.	DWB Taxonomic Checklists, DTN REST API, GBIF and NFDI; ATLAS DER SPINNENTIERE; ARAMOB

Bundesamt für Seeschifffahrt und Hydrographie	The Federal Maritime and Hydrographic Agency (BSH) contributes to use case 5 led by HIFMB. Development of a web-based information network for marine biological data.	CRITTERBASE
Freie Universität Berlin, Botanic Garden and Botanical Museum Berlin		AlgaTerra Information System; **Blast Suche**; **Euro+Med Plantbase**; BioCASe Provider Software BGBM; Data Transformation Service; **GGBN**; Terminology Service; BioCASe Provider Software (Wrapper)
Friedrich-Schiller-Universität Jena		BEXIS2, LAND Portal
German Federation for Biological Data (GFBio e.V.)	GFBio is a not-for-profit association with a special expertise in research data management and a portfolio of services designed along the data lifecycle. Its members are data and service providers from the former German Federation for Biological Data, a DFG-funded project. In NFDI4Biodiversity, GFBio acts as the managing partner (with lead and co-leads in Task Area 5 and its measures) and hosts the HR Pool (now: Data and Software Solutions - DaSS team). GFBio staff has leads and co-leads in several measures and leads three Special Interest Groups for Public Relations, Service Monitoring and Training.	**GFBio Data Search**; Diversity Workbench Management Software; Diversity Workbench Virtual Training Environments at GWDG and SNSB; **GFBio DMP Service**; **Knowledge Base**; **RDM Helpdesk**; **GFBio Portal**; **GFBio Data Submission Service**, Jupyter Notebooks for Training

German Ichthyological Society	GfI-Fische is a natural history society (fish) providing use case 18. It connects the society including monitoring and collection data about fish species with NFDI4BioDiversity.	Biodiversity Warehouse - Fischartenatlas
Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen	The GWDG is a service organization which works in conjunction with the University of Göttingen and the Max Planck Society as a data and IT service center. Furthermore, it carries out independent research in the field of Computer Science. The core areas of research are Cloud Infrastructures, Data Analytics, Data Infrastructures, Research Data Management, High Performance Computing and Scheduling & Resource Management. GWDG contributes its GFBio-related Service Portfolio to NFDI4BioDiversity including the required storage and compute capacities. GWDG will provide its own software developments in this area and beyond (in particular regarding repository solutions) as background.	DWB Management Software; **Sync and Share**; **Storage**; DWB Virtual Training Environments at GWDG and SNSB; DWB Virtual Training Environments at GWDG; Data Repositories
Heidelberg Institute for Theoretical Studies - HITS gGmbH	The Heidelberg Institute for Theoretical Studies was established in 2010 by the physicist and SAP co-founder Klaus Tschira (1940-2015) and the Klaus Tschira Foundation as a private, non-profit research institute. It conducts basic	RightField; SABIO-RK; FAIRDOMHub; FAIRDOM

	research in the natural sciences, mathematics and computer science with a focus on the processing, structuring, and analyzing of large amounts of complex data and the development of computational methods and software. The SDBV group at HITS is a group specialized on scientific data management. It runs the curated data source for reaction kinetics data SABIO-RK, as well as the FAIRDOMHub, a repository for collection and publication of project data, models, SOPs in systems biology and related fields. Within NFDI4Biodiversity, HITS participates in discussions around Task Area 4, and will contribute extensions to the RightField tool for production of ontology-integrate Excel templates.	
Helmholtz Centre for Environmental Research		BioME; LTER-D
Justus Liebig University Gießen		EDGAR, Aruna Object Storage
Leibniz Institute of Freshwater Ecology and Inland Fisheries	IGB is the largest freshwater institute in Germany that leads the international research efforts in the field. It contributes in data science, specifically in the evaluation of heterogeneous, discontinuous, and large data sets, in the implementation and standardization	BenGenDiv, HydrographR

	of comprehensive institutional data management practices, and in establishing a state of the art infrastructure. In Use Case 26, they deployed an R-package that combines the freshwater dimensions of NFDI4Biodiversity and NFDI4Earth, which enables users to link biodiversity data to geographical and limnological features.	
Leibniz Institute of Plant Genetics and Crop Plant Research	The Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) is working on basic and application-oriented research on crop plants. The scientific focus is on the development of new insights into the structure, function and evolution of genetic material, on the conservation, research and development of the hereditary diversity of important cultivated plants, their ancestors and wild relatives. The IPK hosts the Federal Ex situ Gene Bank, representing one of the world's largest germplasm collections, with a current holding of 151,348 accessions. Within the NFDI4Biodiversity project the IPK will provide a FAIR access to various crop plant data from different domains. The IPK is lead for Task Area 3 Measure 1 and Task Area 4 Measure 4.	e!DAL-PGP; MIAPPE; GBIS; EURISCO; BrAPI, Scorpion; DivImpute; MIRA

MARUM – Center for Marine Environmental Sciences, University Bremen	The Centre for Marine Environmental Sciences (MARUM) at University Bremen studies the role of the ocean in the climate system. In collaboration with the Alfred Wegener Institute - Helmholtz Centre for Polar and Marine Research ot operated the Data Publisher for Earth and Environmental Sciences PANGAEA as a permanent facility. UniBremen - MARUM is the applicant organization, including all responsibilities of speaker, and a large Data and Service provider in the NFDI4Biodiversity consortium. UniBremen. MARUM contributes to use cases 3,5 6 and 10, co-leads Task Area 4 and Task Area 5 and leads Task Area 2 Measure 3.	PANGAEA
Museum für Naturkunde – Leibniz Institute for Evolution and Biodiversity Science	Natural history collection and GFBio Data Center for archiving and publishing biodiversity data; MFN is contributing with data curatorial and technical services and expertise. It will contribute to the cooperation with other NFDIs and international political networking activities.	BioCASe Monitor Service (BMS); Naturblick App; Biowikifarm; BioCASe Provider Software MfN; BioCASe Monitor Service software; Animal Sound Archive
Philipps University Marburg		VAT, GeoEngine
Staatliche Archive Bayerns	GDA will contribute their experiences in long-term data preservation, archiving as well as certification. They are experts in data migration, sustained digital readability and the description of data.	Schriftliche Anfragen; Archivfachinformationssystem; Generalisierter XML-Client der Staatlichen Archive Bayerns; Findmitteldatenbank

Bavarian Natural History Collections – Staatliche Naturwissenschaftliche Sammlungen Bayerns	SNSB with 300 employees encompass large zoological, anthropological, paleontological, mineralogical collections, a herbarium (botany and mycology) as well as a biobank (DNA, tissues). The SNSB IT Center is its institutional data center for scientific data in the field of collection and occurrence data, trait data, domain-specific terminologies and taxonomies. Its mission comprises research activities in the field of biodiversity informatics and software engineering (Diversity Workbench). The SNSB will mobilize occurrence and trait data as well as domain-specific/ regional taxon lists. It contributes to use case 4 and 16, is co-lead of Task Area 3, lead of Task Area 3 Measure 3 and involved in several measures as co-lead or contributors.	Bayernflora (Flora of Bavaria); DWB Taxonomic Checklists, DTN REST API, GBIF and NFDI; DiversityMobile App; NaviKey DesktopApp; NaviKey WebApp; MOD-CO A conceptual and procedural metaomics schema; IndExs - Index of Exsiccatae; LIAS - A global information system on lichenized and Non-Lichenized Ascomycetes; DWB Virtual Working Environments at SNSB; DWB Specimen and Occurrence Data, PostgreSQL Cache, BPS API, BMS API, GBIF and NFDI; DWB Management Software; DWB Descriptive Data, DD/DNK API, DiversityNaviKey and NFDI; DEEMY - Determination of Ectomycorrhizae; BioCASe Provider Software installation at SNSB; DWB Virtual Training Environments at GWDG and SNSB
Staatliches Museum für Naturkunde Stuttgart	Natural history collection and GFBio Data Center for archiving and publishing biodiversity data. SMNS is contributing with data curatorial and technical services and expertise. SMNS has strong relations to the natural history society AraGes e.V. and is cooperating as a partner in the AraGes use case 16.	DWB Taxonomic Checklists, DTN REST API, GBIF and NFDI; DWB Management Software

Johann Heinrich von Thünen Institute, Federal Research Institute for Rural Areas, Forestry and Fisheries	Thünen-GDI leads use case 10 which connects NFDI4Agri/FAIRagro and NFDI4Biodiversity by providing biodiversity monitoring as a testbed for data- integration.	**ThünenATLAS**
Universität Bielefeld, Institut für Bioinformatik-Infrastruktur	The Bielefeld Institute for Bioinformatics Infrastructure (BIBI) is an interdisciplinary academic department under the responsibility of the Faculty of Technology at Bielefeld University. The institute covers research and service aspects to handle big data in the life sciences. BIBI contributes to NFDI4Biodiversity in the fields of training and education, bioinformatic capacities (especially cloud computing) and provides access to its national (de.NBI) and international (ELIXIR) network partners. Furthermore it is developing the BIIGLE service for image and video annotation which is to be integrated into the Research Data Commons (RDC) infrastructure and is offering workshops on the service to NFDI4Biodiversity partners.	de.NBI Services Project Support; **BIIGLE**
Leibniz Institut zur Analyse des Biodiversitätswandels (former ZFMK - Zoological Research Museum Alexander Koenig)	Natural history collection and GFBio Data Center for archiving and publishing biodiversity data; the institution is contributing with data curatorial and technical services and expertise. LIB is	DWB Taxonomic Checklists, DTN REST API, GBIF and NFDI; DWB Management Software; BioCASe Provider Software Installation at LIB; Preservation of biodiversity and collection data at LIB;

	leading the GBOL Barcoding initiative and the AMMOD automatic biodiversity monitoring project and will link these major projects with use cases 3 and 4 to NFDI4BioDiversity.	Morph-D-Base; German Barcode of Life Reference Library; ASV-Registry

B. Service Profile Template

Content Responsibility:

Ines Musterfrau

General Information

Service Provider	Link to partner page	
Area of Application	[ ] Planning [ ] Harmonization (e.g. Curation, [ ] Terminology) [ ] Access / Archival [ ] Discover / Exploration [ ] Processing / Visualization [ ] Support / Training [ ] Other: please specify	
Short Description (max 250 words)	Short description (max 250 words)	

Stage of Development	[ ] Development [ ] Pre-alpha [ ] Alpha [ ] Demonstrator [ ] Beta [ ] Release Candidate [ ] Productive [ ] Release *Development: In development without user feedback* *Demonstrator: In development with user feedback* *Productive: Service fully functional, publicly available in a stable release*	
Version	Current Version	
Documentation	Link to Documentation	
License/Terms of Use	Licenses and Link to Terms of Use	
Link to Service/Demonstrator	Link to Service Landing Page or Demonstrator	

Organizational Information

NFDI Service Portfolio	USER-FACING TECHNICAL User-facing: The users interact directly with these services Technical: These services are not visible to the users , but the user-facing services depend on them	
Inclusion in the Service Catalog	[ ] Yes, please : 01 Oct 2020 [ ] No, not right now	
Service provided as part of	[ ] Institutional Mission (Grundaufgabe): *0%* [ ] Project Funding (Projektmittel): *0%* [ ] Others (e.g. Former Project, etc.): *0%* *Required information for the sustainability reporting to DFG*	
Funding	Please provide further information on the funding of the service *Information for the sustainability reporting to DFG*	

Support

Contact	Confluence Profile or Email for the Service Contact. *Who should be contacted for questions about the Service (NOT the* Service Profile*).*	
Helpdesk Contact		Helpdesk

Supported Until	Date => Please type // and choose the date accordingly	Sustainability

Technical Infrastructure

Technical backbone	What is the technical backbone of the service?	
Disaster & recovery plan	Please provide a description or link to a description of the disaster & recovery plan	
Entrance control technical backbone	How is access to the technical backbone managed?	
Stability of operation	Please describe how the continuous operation (24/7) of the service is monitored	

Documentation

Guidance and workflows/templates provided	Is the user provided with some guidance in terms of best practices templates and/or workflows? If so please add the links here	
Communication strategy	How are users informed about news and developments? e.g. Newsletter, Mailing List, Website, etc.	
Registered at:	[ ] Bio.tools : [ ] DFG RIsources: [ ] EOSC Marketplace: [ ] FAIRsharing : [ ] re3data:	Sustain-ability

	[ ] Others: please specify and add link	
Published Information (DOI etc.)	List of publications about the service	

Reporting

Service Category	See Table 1 (until green row)	Monitoring
(Self-) Evaluation	Is some kind of evaluation already available?	Monitoring
KPI Monitoring	Are Key Performance Indicators (KPI) provided for monitoring?	Monitoring
Report Card	Coming soon (will be updated by SIG2 Service, Task Area 3 Measure 1)	Monitoring

Detailed Description

Detailed Description for internal use in evaluation of the Service Profile.

Please include the following points:

- target user group
- skills required
- datatype, -formats, -standards...
- user journey

C. Sustainability Assessment

Criteria for the Assessment of NFDI4Biodiversity Services & Software

Licences/Terms of Use

Please indicate if all services and software are publicly available and operate under an open science/open-source software licence. Please provide a link to the Terms of Use.

Continuity of access

Please state the time until the service or software will be maintained

Organisational infrastructure

Please provide information on the funding structure

Is the service provided as part of your institutional mission (Grundaufgaben) or is it a new service developed within the NFDI framework?

If the funding is mixed, please describe which scope or component is provided as part of the institutional mission and what additional funding is used or needed for.

Technical infrastructure

Please describe how the continuous operation (24/7) of the service is monitored

Documentation

Please indicate if the services and software are published and citable

Feedback/Helpdesk

Please indicate how user support is realised

Criteria for the Assessment of NFDI4Biodiversity Repositories

Licences/Terms of Use

Please indicate if all data is publicly available and the repository operates under an open science licence. Please provide a Link to the Terms of Use.

Continuity of access

Please state the time until the repository will be maintained

Organisational infrastructure

Please provide information on the funding structure

Is the repository provided as part of your institutional mission (Grundaufgaben) or is it a new service developed within the NFDI framework?

If the funding is mixed, please describe which scope or component is provided as part of the institutional mission and what additional funding is used or needed for.

Data integrity, authenticity, and quality

Please describe if documented metadata schemas are in place and used. If possible, please refer to the GFBio Masterlist when listing the standards https://gfbio.biowikifarm.net/wiki/Data_exchange_standards,_protocols_and_formats_relevant_for_the_collection_data_domain_within_the_GFBio_network

Data discovery and reusage

Please state if data can be persistently referred and cited

Technical infrastructure

Please describe how the continuous operation (24/7) of the repository is monitored

Documentation

Please indicate if mission and scope of the repository are published and citable

Feedback/Helpdesk

Please indicate how user support is realised

D. Status of Milestones of Task Area 4

Table D1 - Status of milestones of Task Area 4

Milestone	Milestone Description	Start of work	Target Month	Status
Task Area 4 Measure 1 - Linking data and service providers with the NFDI-RDC				
M4.1.1.	A technical guideline and a reference implementation for the development of standard interfaces by resource providers is available	22	30	In progress

M4.1.2	Workflows for loading and integrating data into RDC established for GFBio data centers and selected use cases	16	24	In progress
M4.1.4	ETL framework for selected data providers implemented and tested	24	48	In progress
Task Area 4 Measure 2 - Data integration & harmonization				
M4.2.1	Storage solution with caching mechanisms for decentralized resources and common metadata exchange and catalogue of core models published for essential subset of use cases	9	24	In progress
M4.2.2	Versioning and synchronisation mechanisms implemented for internal terminologies of the terminology service	13	36	In progress
M4.2.3	Consensus-building process for terminologies and services agreed upon	25	48	In progress
M4.2.4	Annotation, adaptable semantic integration and harmonization tools implemented and services running	25	60	In progress
Task Area 4 Measure 3 - Storage infrastructure				
M4.3.1	Workshop with software developers to adopt the initial design of the REST API. Deployments of basic storage components (e.g. MongoDB, PostgreSQL) in the core storage are available.	04	20	completed
M4.3.2	Prototype implementation of cloud-based storage infrastructure is available. Based on a given list of requirements, a design for the semantic storage exists	04	24	completed
M4.3.3	Working implementations of the semantic storage solutions are available for at least three use-cases	15	36	In progress
M4.3.4	Scalable storage solutions available	04	60	In progress
Task Area 4 Measure 4 - Software application frameworks				

M4.4.2	First version of data discovery portal available	13	24	completed
M4.4.3	First training and support platform established and operational	20	36	In progress
M4.4.4	First version of service discovery portal available	36	48	open
M4.4.5	Selected container-based integrated tools for analysis and visualization available	13	60	In progress
Task Area 4 Measure 5 - Governance & monitoring				
M4.5.2	Rights, obligations and accountabilities for the consortium´s RDC reference implementation sorted and discussed in a series of workshops with NFDI4Biodiversity resource providers (output: report; to be shared with NFDI section Infra)	15	24	In progress
M4.5.3	NFDI4Biodiversity Authentication and Authorization Infrastructure available in place (based on ELIXIR) and evaluated with Use Cases (output: report to be shared with NFDI section Infra)	15	24	In progress
M4.5.4	First set of monitoring and performance measurements for defined KPIs agreed and available for the consortium´s RDC reference implementation	15	30	In progress