A Consortium for the National Research Data Infrastructure (NFDI)

Drawing by Alexa Garin @microbiobac

15.10.2019

contact@nfdi4biodiversity.org

Table of content

DFG

1   General Information

Name of the consortium:

NFDI4BioDiversity: Biodiversity, Ecology & Environmental Data

NFDI4BioDiversität: Biodiversität, Ökologie und Umweltdaten

Summary of the proposal

Biodiversity is more than just the diversity of living species. It includes genetic and phenotypic diversity of organisms, functional diversity, interactions and the diversity of populations and whole ecosystems. The recent Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) report confirmed that our planet experiences a rapid loss of biodiversity with dramatic consequences for mankind. Answers to scientifically and societal relevant questions can only be found based on the availability of data integrated from multiple sources and building on different knowledge domains in science and society. The FAIR data principles provide a framework towards such integration. To foster access to interoperable data, NFDI4BioDiversity focuses on the following six objectives: 1) Promoting research data management as an integral part of biodiversity research, 2) Enabling FAIRness of data, 3) Consolidating FAIRness with quality, 4) Embedding NFDI4BioDiversity into the (inter-)national landscape, 5) Addressing NFDI-wide cross-cutting topics and 6) Promoting collaborative governance, viability and sustainability of data infrastructure services and science. These objectives are implemented by the NFDI4BioDiversity work programme and its five task areas:

Task Area 1 - Community engagement (2involve): This task area will integrate across different knowledge holders and take care of research data management requests related to biodiversity data. A large part of all species data is collected by volunteers in natural history societies, NGOs, museums and citizen science projects as well as collected by different authorities This task area includes 23 use cases with community-driven topics and cross-domain user groups.

Task Area 2 - National and international networking (2connect): This task area ensures that any cross-domain activities and developments of NFDI4BioDiversity are coordinated and in-line with the overall development of the NFDI as well as international activities.

Task Area 3 - Long-term data, tools and service preservation, certification (2consolidate): This task area consolidates existing German infrastructures and data pipelines for long-term FAIR data preservation, archiving and publication and guides them towards certification.

Task Area 4 - Data integration, exploration, and exploitation - the NFDI-Research Data Commons (NFDI-RDC, 4all & 4future): This task area develops a Cloud-based research infrastructure that provides scientists, providers and users with the technical baseline for exchanging data and services in a collaborative research environment. This facilitates the integration, use, and export of biodiversity data as well as the joint development and reuse of services and data.

Task Area 5 - Coordination, collaborative governance & sustainability: This task area implements the governance model and coordination of NFDI4BioDiversity. It includes sustainability aspects and long-term maintenance.

Zusammenfassung in Deutsch

Biodiversität ist mehr als nur die Betrachtung und Erfassung der Vielfalt der Arten auf unserem Planeten. Biodiversität umfasst die genotypische, phänotypische, funktionelle Diversität und Interaktion von Arten, Populationen und Ökosystemen. Der erschütternde Bericht des Weltbiodiversitätsrates (IPBES) hat jüngst bestätigt, dass das Artensterben mit bisher unabsehbaren Konsequenzen für die Menschheit rasant voranschreitet. Die drängenden wissenschaftlichen und gesellschaftlichen Fragen zu den Veränderungen in der Biodiversität können nur durch das Zusammenführen von fundierten Daten und Wissen aus verschiedenen Bereichen beantwortet werden. NFDI4BioDiversity hat sich sechs Projektziele gesetzt, um den Zugang zu dezentralen Daten zu ermöglichen: 1. Forschungsdatenmanagement als integralen Bestandteil der Biodiversitätsforschung fördern, 2. Daten FAIRness ermöglichen, 3. FAIRness um Qualität erweitern, 4. NFDI4BioDiversity in die nationalen und internationalen Aktivitäten integrieren, 5. bei NFDI-weiten Querschnittshemen zusammenarbeiten und 6. die NFDI als Serviceinfrastruktur gemeinsam zu steuern und diese dabei dynamisch und nachhaltig zu gestalten. NFDI4BioDiversity wird diese Ziele in fünf Aufgabenbereichen bearbeiten:

Aufgabenbereich 1 - Einbindung der Biodiversitätsfachkreise (2engage): Dieser Bereich schafft den Schulterschluss zwischen NFDI4BioDiversity und allen Interessensgruppen für Biodiversität und adressiert deren Fragen zum Thema Forschungsdatenmanagement. Eine traditionell herausragende Bedeutung haben dabei die Bürgerwissenschaften, sowie staatliche Behörden im Bereich Umwelt- und Naturschutz. Diese werden mittels 23 z.T. bereichsübergreifenden „use cases" adressiert.

Aufgabenbereich 2 – Nationale und internationale Vernetzung (2connect): Dieser Bereich stellt sicher, dass alle NFDI4BioDiversity Entwicklungen mit den Entwicklungen in der NFDI und auf internationaler Ebene abgestimmt sind, um Insellösungen zu vermeiden.

Aufgabenbereich 3 – Nachhaltige Bereitstellung von Daten, Werkzeugen und Diensten, Zertifizierung (2consolidate): Dieser Bereich dokumentiert und zertifiziert die technische Landschaft der für die Biodiversitätsforschung relevanten Software, Repositorien und Diensten an deutschen Infrastruktureinrichtungen. Ziel ist die Zertifizierung von Repositorien nach den FAIR Prinzipien voranzutreiben.

Aufgabenbereich 4 – Daten integrieren, erforschen und nutzen – die „NFDI-Research Data Commons" (4all & 4future): In diesem Bereich wird eine Cloud-basierte technische Infrastruktur aufgebaut, die es den Forschenden, Datenbereitstellenden und -nutzenden gleichermaßen ermöglicht, Daten kollaborativ zu bearbeiten.

Dies erlaubt sowohl die Integration, Nutzung und Verbreitung von Biodiversitätsdaten als auch die gemeinsame Verbesserung und Nachnutzung von Daten und Services.

Aufgabenbereich 5 – Projektkoordination, gemeinsame Steuerung und Nachhaltigkeit: Dieser Bereich ist für die Koordination und gemeinschaftliche Steuerung von NFDI4BioDiversity im Hinblick auf Nachhaltigkeit und Langfristigkeit der Entwicklungen zuständig.

Applicant institution

Applicant institution	Location
MARUM - Center for Marine Environmental Sciences, University Bremen	Bremen

Name of the consortium spokesperson

Spokesperson	Institution, location
Prof. Dr. Frank Oliver Glöckner	UniBremen-MARUM, Bremen

Co-applicant institutions

Co-applicant institutions	Location
Freie Universität Berlin, Botanischer Garten und Botanisches Museum Berlin (BGBM)	Berlin
Universität Bielefeld, Bielefeld Institute for Bioinformatics Infrastructure (BIBI)	Bielefeld
Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures (DSMZ)	Braunschweig
Leibniz Institute for Farm Animal Biology (FBN)	Dummerstorf
Gesellschaft für Biologische Daten e.V. (GFBio e.V.)	Bremen
Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG)	Göttingen
Heidelberg Institute for Theoretical Studies gGmbH (HITS)	Heidelberg
Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung Gatersleben (IPK)	Gatersleben
Staatliche Naturwissenschaftliche Sammlungen Bayerns (SNSB)	München
Helmholtz Centre for Environmental Research (UFZ)	Leipzig
Justus-Liebig-Universität Gießen (UniGießen)	Giessen
Friedrich-Schiller-Universität Jena (UniJena/iDiv)	Jena
Phillips-Universität Marburg (UniMarburg)	Marburg

Names of co-spokespersons

Co-spokesperson	Institution, location	Task area(s)
Dr. Michael Diepenbroek	UniBremen-MARUM, Bremen	TA 4
Dipl. Inf. Anton Güntsch	BGBM, Berlin	TA 2
Prof. Dr. Jens Stoye	BIBI, Universität Bielefeld	TA 2
Prof. Dr. Jörg Overmann	DSMZ, Braunschweig	TA 5
Prof. Dr. Klaus Wimmers	FBN, Dummerstorf	TA 1
Dr. Ivaylo Kostadinov	GFBio e.V., Bremen	TA 5
Prof. Dr. Ramin Yahyapour	GWDG, Göttingen	TA 2
Dr. Wolfgang Müller	HITS, Heidelberg	TA 4
Dr. Uwe Scholz	IPK, Gatersleben	TA 4
Dr. Dagmar Triebel	SNSB, München	TA 3
Dr. Mark Frenzel	UFZ, Leipzig	TA 1
apl. Prof. Dr. Birgit Gemeinholzer	UniGießen, Giessen	TA 1
Prof. Dr. Alexander Goesmann	UniGießen, Giessen	TA 4
Prof. Dr. Birgitta König-Ries	UniJena/iDiv, Jena	TA 3
Prof. Dr. Aletta Bonn	UniJena/iDiv, Jena	TA 1
Prof. Dr. Bernhard Seeger	UniMarburg, Marburg	TA 4

Participants

Participants	Institution (where applicable), location
Dr. Hubert Höfer	Arachnologischen Gesellschaft e.V. (AraGes-Spinnen), Putbus
Prof. Dr.Stephan Frickenhaus	Alfred Wegener Institute - Helmholtz Center for Polar- and Marine Research Bremerhaven (AWI), Bremerhaven
Dr. Janine Felden	Alfred Wegener Institute - Helmholtz Center for Polar- and Marine Research Bremerhaven (AWI), Bremerhaven
Dr. Jonas Zimmermann	Freie Universität Berlin, Botanic Garden and Botanical Museum Berlin, Research Group Diatoms (BGBM-DI), Berlin
Dr. Gerwin Kasperek	FID Biodiversitätsforschung an der Goethe Universität Frankfurt (BIOfid), Frankfurt
Dr. Gregor von Halem	Bundesamt für Seeschiffahrt und Hydrographie (BSH), Hamburg
Silvia Bender	"Bund für Umwelt und Naturschutz Deutschland e.V. – Friends of the Earth Germany (BUND), Berlin
Prof. Dr. Florian Leese	Aquatische Ökosystemforschung (COST Action), Universität Duisburg-Essen (DNAquaNet), Essen
Dr. Christoph Sudfeldt	Dachverband Deutscher Avifaunisten e.V. (DDA-Vögel), Münster
Klaus-Jürgen Conze	Gesellschaft deutschsprachiger Odonatologen e.V. (GdO-Libellen), Essen
Prof. Dr. Heiko Brunken	Gesellschaft für Ichthyologie e.V. (GfI-Fische), Solingen

Prof. Dr. Volkmar Wolters	Gesellschaft für Ökologie e.V. (GfÖ), Berlin
Prof. Dr. Helmut Hillebrand	Helmholtz Institute for Functional Marine Biodiversity at the University of Oldenburg (HIFMB), Oldenburg
Stephan Karger	Hessisches Landesamt für Natur, Umwelt und Geologie (HLNUG), Wiesbaden
Dr. Jana Moser	Leibniz Institut für Länderkunde (IfL), Leipzig
Prof. Dr. Hans-Peter Grossert	Leibniz-Institut für Gewässerökologie und Binnenfischerei (IGB), Berlin
Prof. Dr. Adrian Paschke	Institut für Angewandte Informatik (InfAI), Leipzig
Dr. Gotthard Meinel	Leibniz-Institut für ökologische Raumentwicklung (IÖR), Dresden
Dr. Holger Beer	Julius-Kühn-Institut Bundesforschungsinstitut für Kulturpflanzen (JKI), Quedlinburg
Dr. Kai Gedeon	Landesamt für Umweltschutz Sachsen-Anhalt (LAU),
Holger Lueg	Sächsisches Landesamt für Umwelt, Landwirtschaft und Geologie (LfULG), Dresden
Dr. Jana Hoffmann	Museum für Naturkunde Berlin (MfN), Berlin
Dr. Miguel D. Mahecha	Max-Planck-Institut für Biogeochemie (MPI-Biogeochem), Jena
Stefan Munzinger	Naturgucker gemeinnütziger e.G. (Naturgucker), Northeim
Dr. Steffen Caspari	Netzwerk Phytodiversität Deutschland (NetPhyD-Pflanzen), Schiffweiler
Prof. Dr. Jörg Müller	Nationalparkverwaltung Bayerischer Wald (Behörde im Geschäftsbereich des Bayerischen Staatsministeriums für Umwelt und Verbraucherschutz) (NP Bayerischer Wald), Grafenau
Jan Rommelfanger	Nationalparkamt Hunsrück-Hochwald (NP Hunsrück-Hochwald), Birkenfeld
Dr. Margit Ksoll-Marcon	Staatliche Archive Bayerns (SAB), München
Prof. Dr. Thomas Hickler	Senckenberg Gesellschaft für Naturforschung (SGN), Frankfurt am Main
Prof. Dr. Johanna Eder	Staatliches Museum für Naturkunde Stuttgart (SMNS), Stuttgart
Prof. Dr. Wolfram Horstmann	Niedersächsische Staats- und Universitätsbibliothek Göttingen (SUB), Göttingen
Florian Hoedt	Thünen Institute (Thünen-GDI), Braunschweig
Prof. Dr. Thomas Kneib	Universität Göttingen (UniGöttingen), Göttingen
Prof. Dr. Christian Wirth	Universität Leipzig (UniLeipzig), Leipzig
Dr. Peter Grobe	Zoologisches Forschungsmuseum Alexander Koenig (ZFMK), Bonn
Dr. Nicolas Dittert	Leibniz Zentrum für Marine Tropenforschung (ZMT), Bremen

Names and numbers of the DFG review boards (DFG Fachkollegien) that reflect the subject orientation of the proposed consortium

201   Basic Research in Biology and Medicine

202   Plant Sciences

203   Zoology

204   Microbiology, Virology and Immunology

207   Agriculture, Forestry and Veterinary Medicine

2   Consortium

2.1   Research domains or research methods addressed by the consortium, objectives

Biodiversity is more than just the diversity of living species. It includes genetic and phenotypic diversity of organisms, functional diversity, interactions and the diversity of populations and whole ecosystems. Mankind continues to dramatically impact the earth's ecosystem which is the foundation of human well-being [1]. A general understanding of the status, trends, and drivers of biodiversity on earth is urgently needed to determine management options and devise conservation responses. Answers to scientifically and societal relevant questions can only be found through the availability of data integrated from multiple sources and building on different knowledge domains in science and society. This needs to be realized following the FAIR (Findable, Accessible, Interoperable, Re-usable) data principles [2]. To foster easy access to interoperable data, NFDI4BioDiversity focuses on the following objectives:

Promoting research data management as an integral part of biodiversity research:

In an ideal world research data management (RDM) is seamlessly embedded within the complete scientific workflow. The insufficient co-design between biodiversity research and digitalisation has contributed to the fact that large parts of the current data management landscape is fragmented, ephemeral, not efficient, and lacks quality. Within Task Area 1, NFDI4BioDiversity will ensure that RDM becomes an integral, funded and credited part of biodiversity research, which is a prerequisite for successful data science approaches and a major requirement for the NFDI in general.

Enabling FAIRness of data:

As professional data management is still almost absent or underdeveloped in science, it is often done according to individual, not necessarily FAIR compliant, workflows. Data for biodiversity research range from biolab experimental studies to automatic device recording in the field. The data generation is often done by students and young researchers without much data management experience. The bandwidth of scientific questions and approaches in biodiversity and ecological communities aggravates the situation, also with respect to data sources and formats.

As a speciality in biodiversity research a considerable amount of data is collected by citizen scientists and (semi-)professionals, which may not even have their data digitized. Reflected in our use cases and further addressed by Task Areas 1, and 4, NFDI4BioDiversity will support structured digital data acquisition, remove technical hurdles in data mobilisation, eliminate deficiencies, inconsistencies, and incompatibilities in data structures and semantics which impede effective interoperability and re-use of data. It will support education and training events to raise awareness and anchor the requirements in the biodiversity community. In Task Area 2 NFDI4BioDiversity will contribute to develop common minimum requirements standards across NFDI.

Consolidating FAIRness with quality:

Data of ambiguous quality will constrain the trust and acceptance of FAIRness. Since "quality of data" strongly depends on the application, the provenance of data must be conserved in the archiving and publication procedure. Data curation ensuring that data transformations are adequately documented in the metadata is therefore key for any re-use of data. Explicit definition of terms, formats and the use of controlled vocabularies and ontologies as well as clear licensing information must be ensured. In Task Area 3 we will set up measures to evaluate and prioritise repositories where quality assurance ranging from data provenance to data integrity for long-term archival is implemented. Certification will be an important measure to gain trust in the community with respect to data deposition and re-use.

Embedding NFDI4BioDiversity into the national & international landscape:

As biodiversity science is not limited to national boundaries integration of any future services into the existing international service landscape is crucial. Data and information must be exchanged independently of local, national or international borders. The partners of NFDI4BioDiversity have a strong standing in international biodiversity and data activities like GBIF [3], CETAF [4], EUBON [5], GEO BON [6], TDWG [7], (e)LTER [8, 9], NEON [10], ECPGR [11], RDA [12], GOOS [13], FREYA [14], FAANG [15], ELIXIR [16], ENVRI [17], EUDAT [18], DataONE [19], EOSC-Hub [20]. Within Task Area 2 we will team up to coordinate international implementations and harmonise access to existing services. Structuring and leading the biodiversity community in Germany will boost the standing of Germany in generic initiatives like the European Open Science Cloud (EOSC) [21].

NFDI-wide cross-cutting topics:

NFDI4BioDiversity has initiated an NFDI-wide discussion about cross-cutting topics. As a first result 21 NFDI consortia consolidated and agreed on a set of topics as outlined in the "Berlin Declaration on NFDI Cross-Cutting Topics" [22]. By signing the declaration each NFDI consortia commits to collaboratively work on these topics.

As agreed, leadership for certain topics will not be claimed at this time but addressed in inter-consortium working groups as soon as the funding decision has been made. In agreement with the Berlin Declaration, NFDI4BioDiversity has a special interest in the following topics: Collaborative governance, community involvement with a special focus on graduate education and cultural change, research data commons, terminology management and metadata harmonisation as well as quality assurance and certification.

Promoting Collaborative Governance, Viability and Sustainability

To govern and sustain NFDI4BioDiversity, the not-for-profit association GFBio e.V. will be instrumental as the central coordination and management unit of NFDI4BioDiversity as outlined in the section 2.5 and implemented in Task Area 5. The speaker together with GFBio e.V. will be the prime contact point for all data providers, users, co-applicants and participants as well as the DFG, Board of Governors, NFDI Directorate and Senate. Direct shaping and governing the development of GFBio e.V. is realised by a membership in GFBio e.V. To professionally operate NFDI4BioDiversity, GFBio e.V. will employ a set of full-time coordinators to implement and monitor the progress of the work program. Viability of the consortium will be guaranteed by evaluation and admission procedures as well as EXIT-strategies for co-applicants and participants. Sustainability of all data is granted by the institutionally funded repositories. Modular software development and a well-documented, open-access codebase will second its long-term maintenance. To realise a collaborative governance across NFDIs, NFDI4BioDiversity is open for any cooperative development to realise a single legal entity e.g. an NFDI e.V.

2.2 Composition of the consortium and its embedding in the community of interest

The fundamental equilibrium between abiotic and biotic diversity and anthropogenic influences is key for sustainability. However, it is difficult to measure biodiversity changes on different scales across time and space as information remains scattered across databases with varying qualities, standards, semantics, formats and access schema. With its highly committed partners, embedded in an international network, NFDI4BioDiversity will provide its users and data providers with data products, information, knowledge, training, tools and services for the management of research data. At the time of writing the consortium is composed of the applicant institution, 14 co-applicants and 35 participants. The combined expertise ranges from microbial to plant and animal biodiversity and includes seven natural science collections for bio- and geodiversity data. The datacenters at the natural science collections (BGBM, DSMZ, SNSB, MFN, SGN, SMNS, ZFMK) include the largest German natural history research collections, taxonomic facilities, the network of botanical gardens and the world's most diverse microbiological resource collections, which together not only host more than 75% of all museum objects (150 millions) in Germany and >80% of all described microbial species, but also represent the biggest and internationally most relevant data repositories.

Additionally, the IPK Genbank in Gatersleben holds one of the largest collections of crop plants and their wild relatives and the Leibniz Institute for farm animal biology provides expertise to understand and use animal functional biodiversity for livestock farming. Biodiversity related environmental data is supplied by PANGAEA. Dedicated co-applicants will provide the technical infrastructure for NFDI4BioDiversity ranging from storage to cloud-based compute power and participation in international standard developments for data linkage and networking almost in-kind.

Finally, NFDI4BioDiversity builds on the established, DFG-funded German Federation for Biological Data (GFBio) project. GFBio consists of 20 institutions in Germany, with many of them are the pillars of NFDI4BioDiversity. GFBio encompasses technical, organizational, financial, and community aspects to raise awareness for research data management in biodiversity research and environmental sciences. To foster sustainability and preserve expertise across this federated infrastructure, the not-for-profit association "Gesellschaft für biologische Daten e.V. (GFBio e.V.)" has been installed in 2016 as an independent legal entity. The Speaker together with GFBio e.V. will coordinate NFDI4BioDiversity to safeguard experiences and expertise gained in the GFBio project.

Besides the traditional and well-established biodiversity facilities, NFDI4BioDiversity has taken care to truly represent the citizen science and volunteer community which is of utmost importance in biodiversity research. To pick out one example, the naturgucker.de portal [23] is the largest information system for volunteer nature watchers in Germany. naturgucker.de currently includes more than 1.5 million natural images of almost 38,000 species from more than 46,000 areas worldwide. In addition, they store 10.5 million observations of plants, animals and fungi worldwide. Nearly 68,000 volunteer nature lovers participate in the social observation network naturgucker.de and contributed more than 550,000 observations and images of nature. Furthermore, NFDI4BioDiversity is proud to welcome a range of volunteer associations connected to spiders, plants, fishes, birds, dragonflies as well as natural park authorities, federal state environmental agencies and Friends of the Earth Germany to provide data, expertise and access to their communities. Detailed information of each co-applicant and the participants as well as their contributions to the work programme are given in Table 2.2.1: Summary of NFDI4BioDiversity's participants, their roles and contributions and Table 2.6.1: Expertise, resources and infrastructures of the co-applicants and participants available for the NFDI.

Analysis of the existing research and infrastructure environment

Biological sciences and the investigation of the living world around us trace back to ancient times, since mapping and understanding the fauna and flora was crucial for survival and is now the basis for feeding 7.5 billion people. Recent reports on the extinction of species, habitats or whole ecosystems have further intensified biodiversity research.

The tight interaction with our user communities has shown that the biodiversity research environment is highly dynamic. Research is done on all size scales ranging from microbes to whales and in any habitats from the deep biosphere to the stratosphere and above. Biodiversity data is diverse: originating from spot measurements, time series of observation data to molecular (sequence) data, pictures, videos and sound files up to specimen data in natural history collections. Biodiversity data is scattered: besides data from research projects stored in a multitude of (local) databases and repositories, highly relevant data are hosted by state and federal authorities, nature conservation organisations and non-governmental organisations.

The biodiversity community relies on an infrastructure environment consisting of a high number of almost incompatible (sub-) domain and specific data repositories to store their data. Re3data lists 816 data repositories flagged with "biology" [24]. Consequently, just findability and accessibility of the data is constrained or even impossible. The lack of documentation including license information intensifies the problem. Most of the biodiversity data repositories are set up on proprietary management systems, which prohibits data exchange. These "data and information silos" [25] are typical for research groups, but also for governmental and non-governmental nature conservation and biological science organisations. Within the GFBio project we started to open these data silos to transform at least some of the data to well-defined, community standard-compliant data products. First persistent data mobilisation pipelines were established to create and publish highly relevant data products for research. In parallel they serve for several national and international portals, the GFBio VAT tool, GBIF and a couple of domain-specific platforms of which many are partners in NFDI4BioDiversity [26].

In summary the current research and infrastructure environment hinders the interoperability and re-usability of data to answer scientifically and socially relevant questions.

Means of communication with our providers and users, sensing the specific needs of our community

NFDI4BioDiversity builds on more than 20 years of experience in providing information systems and infrastructures for providers and users. As exemplified below the applicant and co-applicants have a strong standing in national and international communities which will be instrumental for the trust, acceptance and success of NFDI4BioDiversity and NFDI.

The GFBio community

As part of the GFBio project an intensive exchange with the biodiversity community has been practised. More than 30 GFBio roadshows in over 20 locations in Germany were organised, reaching more than 450 individual researchers as multiplicators for intense discussions. Further outreach activities included expert round tables, training events, and participation in targeted workshops and conferences. This was complemented by the distribution of promotional material and social media activities.

Additionally, a survey on the users' perspectives on research data management was conducted and disseminated through various communication channels to reach scientists from a broad range of subdisciplines within the field of biological and environmental sciences [27]. The goal of these activities was to raise awareness for the importance of research data management as well as to collect the community requirements. The results were the basis for an informed planning of the data management support and efficient user engagement strategies in NFDI4BioDiversity.

GFBio's professional helpdesk and ticket system with first and second level support teams provide another instant line to the needs of our community. More than 150 user requests have been registered in GFBio's helpdesk's ticket system since 2016, split into assistance in the preparation of data management plans and support in research data management. As a result, a considerable amount of data was already submitted for long-term archival and publication. GFBio has mediated the archiving of more than 750 environmental data sets at PANGAEA, each now publicly available and citable via DOI. More than 80 biodiversity data sets, consisting of over 14 million persistently identifiable single data units, such as species occurrences with multimedia files, were deposited, quality-controlled and archived in the collection data centers and are already publicly available. More than 60 sequence data submissions were done via the dedicated molecular workflow of the GFBio submission system. These ~900 data sets have already been cited over 1300 times.

In summary, the experience with GFBio has shown that the community is willing to accept and trust a fully commissioned data infrastructure for research data management. It also shows that our principles and strategies of user and community engagement to provide and make use of curated high-quality data are successful.

The German Natural Science Research Collections community

The six major German Natural Science Research Collections BGBM, MfN, SGN, SMNS, SNSB, ZFMK and the German culture collection DSMZ were transitioned from GFBio and are co-applicants or participants in NFDI4BioDiversity as well as members of GFBio e.V. They are also active in the consortium "Deutsche Naturwissenschaftliche Forschungssammlungen" (DNFS, see letter of support 6) [28], as well as in the "Consortium of European Taxonomic Facilities" (CETAF) [4]. Taken together they store more than 150 million physical objects, most of them biological objects relevant and available for biodiversity research purposes.

The more than 1000 researchers, curators, taxonomists and data managers located at the collections provide a wide range of services for the users starting with curation for expanding the collections to training students, organising exhibitions and undertaking research studies. They play a key role in linking heterogeneous user communities, e.g. (local) volunteer taxon experts with academic researchers in molecular phylogeny and systematics, ecology, evolutionary biology and biologists working at regional and national nature conservation agencies.

Taxonomists employed at natural science collections are traditionally often members of standardisation initiatives (object preservation, data etc.) and of committees to contribute in international legal codes, e.g. for nomenclature like ICN (International Code of Nomenclature for algae, fungi, and plants) [29], ICZN (International Commission on Zoological Nomenclature) [30] and ICNP (International Code of Nomenclature of Prokaryotes) [31].

The collections as material repositories of NFDI4BioDiversity are committed to sense and support the biodiversity community in the deposition of new materials and host data related to objects in their data repositories.

GBIF – the Global Biodiversity Information Facility network and community

GBIF – the Global Biodiversity Information Facility – is an international network and research infrastructure funded by the world's governments and aimed at providing anyone, anywhere, open access to data about all types of life on Earth [32]. The main objective of GBIF Germany [33] is the focussed collection and mobilisation of data within Germany's research community and natural history collections as well as the available observation data. All GBIF-D node institutions, i.e. BGBM, DSMZ, MfN, SGN, SNSB, ZFMK are either co-applicants or participants within the NFDI4BioDiversity consortium and contribute through their long-term experience with data standards like DwC (Darwin Core) and ABCD (Access to Biological Collection Data) and operational technical infrastructure for management, transforming and publishing of biodiversity data. They serve as mediators and "brokers" between biodiversity researchers to make their data available through GBIF. Building on the GBIF facility and community the NFDI4BioDiversity infrastructure will organize the mobilisation, standardisation, archiving and publication of various types of data products as delivered by a wide array of biodiversity research studies including citizen science monitoring projects.

The German Network for Bioinformatics Infrastructure (de.NBI) community

The German Network for Bioinformatics Infrastructure (de.NBI) [34] is a dynamic and distributed bioinformatics infrastructure, which started in March 2015 as an academic funding initiative of the German Ministry of Research and Education (BMBF). At present, the de.NBI network consists of 42 projects organized in eight service centers. The activities of the service centers are clearly distinct. Three centers have an organismic background and deal with microbes, plants and human. Three centers focus on RNA bioinformatics, proteome bioinformatics, and integrative bioinformatics. The remaining two centers deal with databases and data management. The main goal of de.NBI is to provide first class bioinformatics services and training for life sciences research in academia and industry. Since 2015 de.NBI has organized more than 200 training courses with over 4500 participants. Through these efforts, de.NBI has been very successful in bringing together scientific excellence in the field of bioinformatics and fostering collaboration between researchers and institutes within Germany.

By partnering with de.NBI, NFDI4BioDiversity makes full use of this large- and well-established network of service providers to attract the users with analysis capacities to subsequently provide support in research data management.

PANGAEA – World Data Center and Data Publisher for Earth and Environmental Science

UniBremen-MARUM in collaboration with the AWI, operates the Data Publisher for Earth and Environmental Science PANGAEA [35] as a permanent facility. Founded in 1992, PANGAEA has demonstrated its long-term perspective by a certification of the ICSU World Data System and the CoreTrustSeal, and is accredited by the WMO as Data Collection and Processing Center (DCPC). The system is operated in compliance with the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. So far PANGAEA has participated in more than 350 national, EU, and International projects, see [36]. More than 15,000 users have registered at PANGAEA (annual growth about 1,100 users). PANGAEA is internationally accepted by serving around 15,000 unique users per month. With every submission the PANGAEA curators are in close contact with the data providers to not only receive and curate their data, but to sense their needs and specific requirements. With UniBremen-MARUM being the applicant and AWI being participant, NFDI4BioDiversity will have a direct connection to this large user community.

Preparing NFDI and NFDI4BioDiversity

In preparation of the NFDI, the life-science community organised itself under the NFDI4Life-Umbrella initiative, consisting of NFDI4BioDiversity, NFDI4Health, NFDI4Microbiota, NFDI4Neuroscience, NFDIAIRR, NFDI4Agri and the German Genome-Phenome Archive (GHGA). Over the last two years multi-lateral discussions and workshops have been organised to gather a common understanding of research data management across the different domains. Although NFDI4Life will not hand in an application at this point, the knowledge and experiences gained have shaped our understanding of the needs and challenges of other user consortia and communities and how they can be addressed in a collaborative effort.

As described in section 2.3 we are in close contact with many other consortia and have attended some of their consortia meeting. A better mutual understanding of the user and provider needs has already been reached and concrete cooperation between task areas as well as cross-cutting topics have been filed in the "Berlin Declaration on NFDI Cross-Cutting Topics".

NFDI4BioDiversity Community Workshop, prioritisation of work

In June 2019 we organised a community workshop to gather the voice of the community and their needs with respect to NFDI4BioDiversity as a potential consortium for NFDI. Fifty representatives from universities, research institutes, learned societies, associations, authorities and NGOs answered our open call for participation. They outlined their expectations and potential contributions to NFDI4BioDiversity in 14 short talks.

In summary, all participants agreed that NFDI4BioDiversity has the potential to harmonise the community, guide a common understanding of metadata standards and foster interoperability and exchange of data. Initiated by the discussions, 23 community driven use cases have been designed which are now part of NFDI4BioDiversity Task Area 1 and summarized in Table 4.2.1. Based on the demand of the community, we will establish education and training and a cross-domain graduate education programme for research data management and data science in Bremen as a blueprint

for future NFDI programs. The community workshop in conjunction with our long-time experiences in working with the biodiversity community on many angles were the foundation for creating the task areas and measures in our work program.

How do we interact and involve the providers and users in the development of NFDI4BioDiversity?

Based on the consortium's solid experiences with users and providers we would like to state that a strict distinction between data providers and users does not reflect reality. Users as we define them, hold in most cases both roles as data producers and consumers by conducting their experiments and measurement as well as collecting and integrating data from different sources for synthesis and analysis. Exceptions might be governmental agencies and authorities, who are obliged to run monitoring programs and provide data for political decision making and research. In parts this is also true for data that is collected by volunteers e.g. in citizen science campaigns, since they often do not have the capacities to perform integrated data analysis. Therefore, the measures in Task Area 1 on user engagement is fine-tuned towards the needs of the different groups and stakeholders.

Citizen Scientists, associations, learned Societies and NGOs

Citizen Science has always played an important role in the biodiversity domain dating back to the 18th century [37]. In fact, 80-90% of species occurrence and trait data is collected by experts and volunteers in natural history associations and societies, NGOs, taxonomic facilities like museums and citizen science projects [38, 39]. An important development is that the extinction or loss of species has recently gained a huge societal and political impact. To give an example: In Bavaria citizens expressed strong concerns against the government and changed the nature conservation act towards better protection of biodiversity [40]. The high awareness to protect our living world is echoed by the willingness of the public to further improve the data coverage as the basis for research and decision making. Consequently, NFDI4BioDiversity has set up measures 1 and 4 in Task Area 1 including specific community-driven use cases to allow a strong involvement of these stakeholders. By working with them over many years, we have understood that visibility and credibility of their valuable work are of utmost importance.

Providing state of the art digital infrastructures for data acquisition and dissemination, ranging from Apps to webpages, is still a major issue for this group. For the design of solutions that really help, NFDI4BioDiversity has an IT task force team in place to assist them in data acquisition, standardisation, interoperability as well as IT and software development. We will develop a co-design process to provide custom-tailored - personalised - solutions that fit their needs and in parallel mobilises data for re-use.

Government agencies and authorities

Several attempts of the co-applicants have been undertaken over the last years to mobilise data for public re-use. Besides legal issues, it is mostly a lack of workforce to transform their data and make them accessible via machine-machine communication. Since legal questions can be expected to show up on several occasions in NFDI, they were defined as one of the cross-cutting topics where we expect to find solutions across NFDI in a cooperative way. To better involve authorities and support the mobilisation of their data we will enter a negotiation process to provide targeted financial support for on-site data managers via an application procedure or send out our task force team to help with technical expertise if needed. By having a couple of federal agencies and authorities as supporters (see letters of support 2 & 14) and leading participants in the use cases on board, they can act as our bridgeheads and showcases to trigger this process.

Researchers

Researchers are from our experience the most complex target group. In general, they are highly interested in using structured and harmonized datasets for their research questions. Providing data is in most cases the main constraint since in the current research system the personal reputation and career path are highly correlated with scholarly publications. Making data public at an early stage of the data life-cycle raises the fear that others might be faster in publishing the next high-ranking paper, impairing their professional career [41]. The complexity of the current scientific credit and reputation system cannot be resolved by NFDI4BioDiversity alone. Initiating a cultural change has already been identified as a major cross-cutting topic by all consortia. On the positive side, the open science and open access movement propagated by funders on the national and the EU level, has raised awareness that data is not a private good, but part of the public heritage [42]. As digitalisation proceeds more and more institutes and universities develop binding research data management guidelines following the FAIR data principles (see letter of support 9). Although investment in research data management is in many cases not immediately rewarding, it is commonly accepted that the public availability of a rich set of high-quality data will allow new kinds of statistical and machine learning approaches and further science and innovation. Consequently, many researchers are now developing an intrinsic motivation for sharing their data.

As part of the GFBio project, we are in contact with a network of data managers at universities and institutes that act as the on-site contact point for any questions in research data management (front-office). NFDI4BioDiversity acts in the back-office, taking over and dispatching the questions to the right partners as well as providing training and (cross-cutting) educational activities as part of Task Area 1 and 2. We will further incentivise data management by providing tools and platforms for data acquisition in Task Area 1 and integrated added value services as part of the research data commons (NFDI-RDC) in Task Area 4.

In general

The composition of the consortium is well balanced with respect to 1. the role, expertise, resources and capacities of the co-applicants in the national and international context. Our participants add an excellent standing in the community and provide special expertise to round-up the portfolio, 2. the completeness of partners with respect to the objectives and work program and 3. the ratio of infrastructures and universities with a strong standing in education and research (Figure 2.2.1). Taken together, we are convinced that the consortium itself is an incentive for users and providers to participate in the NFDI process. Last but not least, the direct involvement of institutes and users in the decision-making of NFDI4BioDiversity can be practised by a membership in GFBio e.V.

Applicant, co-applicants and participants and their roles with respect to the work programme
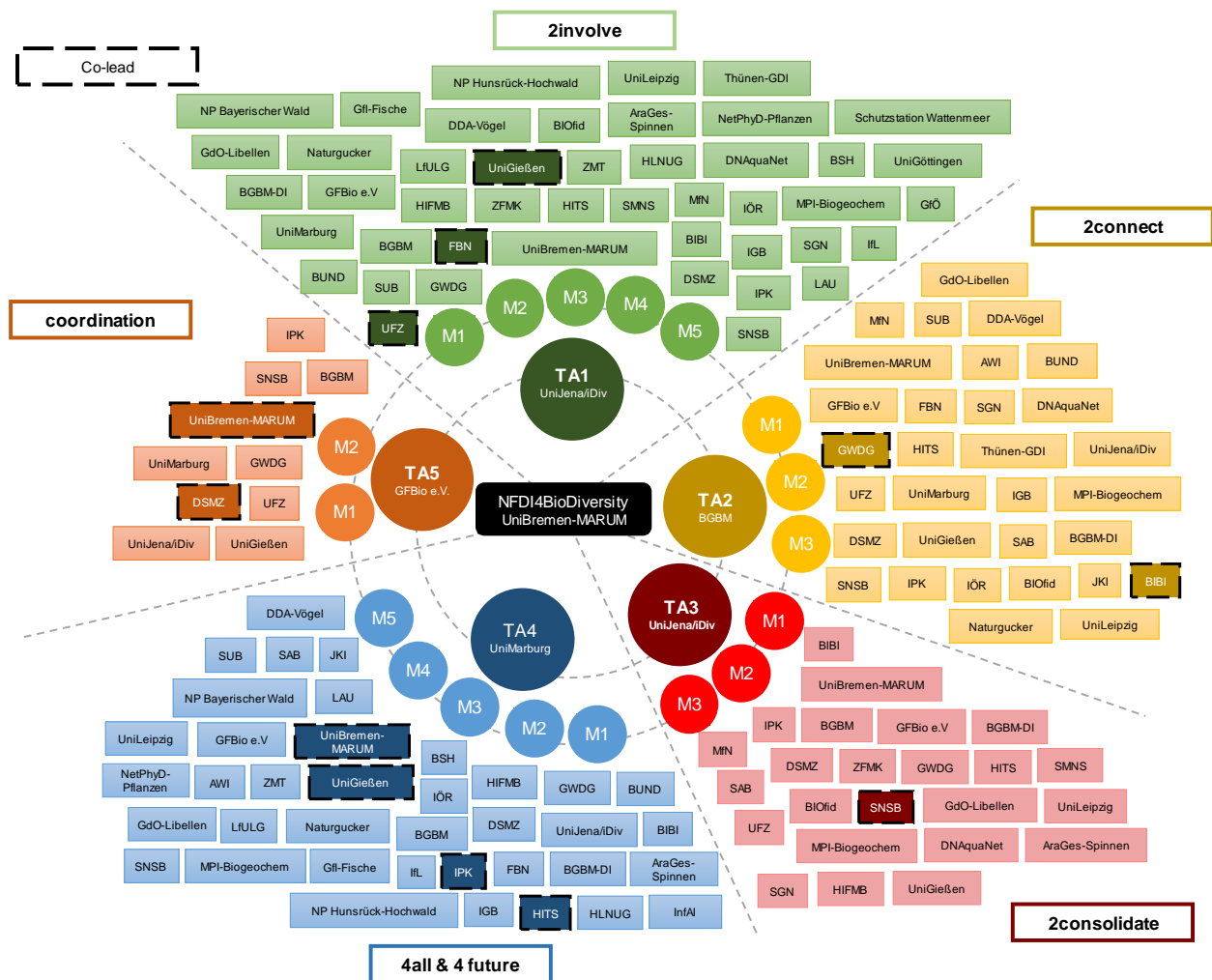
Figure 2.2.1: Graphical overview of application, co-applicants and participants and their contributions to the task areas and measures

UniBremen-MARUM - Center for Marine Environmental Sciences, University Bremen (applicant, Speaker)

The Centre for Marine Environmental Sciences (MARUM) at University Bremen focuses on the role of the ocean in the climate system. MARUM is the home of the DFG Cluster of Excellence „The Ocean Floor – Earth's Uncharted Interface" as well as national and international research projects. MARUM in collaboration with the Alfred Wegener Institute - Helmholtz Center for Polar and Marine Research (AWI), operates the Data Publisher for Earth and Environmental Science PANGAEA as a permanent facility. PANGAEA offers access to >380,000 datasets (annual growth about 11,000 datasets) comprising of about 16 billion individual observations or measurements. UniBremen-MARUM coordinates the German Federation for Biological Data (GFBio) for seven years and leads the BioData service center for biological data and the special interest group 2 on service and service monitoring in the German Network for Bioinformatics Infrastructure de.NBI.

Role: UniBremen-MARUM is the Speaker of NFDI4BioDiversity. Furthermore, it will contribute to scientific data management, long-term data archiving, data publication, data dissemination and interoperability with national, European and international data information systems. Besides providing access to PANGAEA's large user community UniBremen-MARUM will support certification and user-tailored development of information infrastructures. UniBremen-MARUM will contribute to use case 3, 5, 6, and 10. It co-leads Task Areas 4 and 5 and contributes to Task Areas 1 M1-3, 2 M3 and 3 M3.

BGBM - Freie Universität Berlin, Botanischer Garten und Botanisches Museum Berlin (co-applicant)

The Botanic Garden and Botanical Museum Berlin (BGBM) of the Freie Universitat Berlin is a centre of biodiversity research in Europe, housing extensive scientific collections of herbarium specimens (about 3.5 million), one of the world's largest living plants collections, as well as the most complete botanical library in Germany. The focal point of R&D activities is taxonomic information systems, networking of distributed primary biodiversity information and research workflows. The BGBM hosts a wide range of databases and information systems and plays also an active role in international organisations such as GBIF, CETAF, TDWG and DiSSCo.

Role: BGBM will continue as a data centre for archiving and publishing research data within the framework of the NFDI4BioDiversity infrastructure. It will provide its network and expertise on national and international networking, as well as its terminology services and LOD. BGBM will lead Task Area 2 and contribute to Task Areas 1 M3&5, 3 M2&3, 4 M1&2 and 5 M1.

BIBI - Bielefeld Institute for Bioinformatics Infrastructure at Bielefeld University (co-applicant)

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. It contributes to the BMBF funded "German Network for Bioinformatics Infrastructure" (de.NBI) to provide comprehensive state-of-the-art bioinformatics services and training to users in basic and applied life sciences research from academia and industry and fosters cooperation of the German bioinformatics community with international bioinformatics network structures such as ELIXIR. The coordination and administration office of the de.NBI-network and the national ELIXIR node ELIXIR-Germany are located at the BIBI.

Role: BIBI will contribute to NFDI4BioDiversity in the field of training and education, bioinformatic capacities, provide its expertise in national and international networking, especially with respect to Europe and ELIXIR. Of special importance for the NFDI4BioDiversity project is the free access to the federated de.NBI cloud. BIBI will co-lead Task Area 2 and contribute to Task Areas 1 M2, 4&5, 3 M1&2 and 4 M3-5.

DSMZ – Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures (co-applicant)

The Leibniz-Institute DSMZ is one of the largest microbiological resource centers worldwide and hosts more than 70,000 biological resources, including 30,000 different bacterial and 5,000 fungal strains. DSMZ operates a large-scale sequencing facility, server structure, and bioinformatics pipelines. It has long-term expertise in bacterial genomics and is a partner of the Genomic Encyclopedia of Bacteria and Archaea (GEBA). The DSMZ hosts The Bacterial Diversity Metadatabase (BacDive), the largest database for standardized bacterial phenotypic data. A legal department on CBD and Nagoya has been recently installed.

Role: The DSMZ contributes its long-standing experience in curation of microbial resources and data. With BacDive it will provide a microbial repository already offering information for more than 80% of the so far described prokaryotic diversity. Moreover, the DSMZ provides the prokaryotic taxonomy terminology service Prokaryotic Nomenclature Up-to-date (PNU). DSMZ contributes to the consolidation of existing services, interoperability between data and service providers and the harmonisation of metadata and terminology services. It will provide advice on legal questions and certification. DSMZ co-leads Task Area 5 and contributes to Task Areas 1 M2&3, 2 M1&3, 3 M3 and 4 M1&2.

FBN – Leibniz Institute for Farm Animal Biology (co-applicant)

The Leibniz Institute for Farm Animal Biology (FBN) investigates the intrinsic biological processes of animals in their specific environmental contexts to understand and use their functional biodiversity to develop solutions for sustainable livestock farming. Globally, animal breeding led to a biodiversity that is reflected in the multitude of livestock breeds, some of which are endangered. FBN studies the pathways and abilities that animals developed during evolution and the biological changes that farmed animals underwent in the course of domestication and breeding to manifest traits and adapt to their environments. The FBN generates a tremendous quantity of biological data from a vast variety of farm animals (genomics and epigenomics, transcriptomics, proteomics, metabolomics, microbiomics and ethologics).

Role: Quality assured research data are the basis for scientific knowledge and further research. FBN will contribute its comprehensive experience in handling large amounts of data of different data types, and in analysing, modelling and visualisation of these datasets. FBN will co-lead Task Area 1 and contribute to Task Areas 2 M2&3 and 4 M2.

GFBio e.V. – German Federation for Biological Data e.V. (co-applicant)

GFBio e.V. was founded in 2016 as an independent, not-for-profit association dedicated to RDM. At the time of writing the association has 33 members (18 institutional, 15 private) ranging from collections to computing centers and research institutes interested in participating in the latest RDM developments [43].

GFBio e.V.'s main objectives are building and supporting an active network of experts and infrastructures as well as providing a single point of contact for a rich portfolio of services, covering the whole data life-cycle, from data management plans to data publication.

Role: The Speaker together with GFBio e.V. will coordinate NFDI4BioDiversity and represent NFDI4BioDiversity by acting as the single contact point for all data providers, and users, stakeholders and shareholders as well as the NFDI board and panels. It will coordinate the co-applicants, participants and users/providers as well as the helpdesk, service portfolio, public relations and training/education activities. It will host a task force for IT and software development. GFBio e.V. leads Task Area 5 and contributes to Task Areas 1 to 4 on all measures.

GWDG - Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (co-applicant)

GWDG is an IT service provider and data centre for the Max Planck Society and the Georg-August-University Göttingen. GWDG provides specific research-oriented solutions for data management, data analytics, cloud computing, and distributed infrastructures. GWDG also operates large-scale Cloud services, HPC systems, and data repositories.

Role: GWDG has the role of a service provider and integrator within NFDI4BioDiversity. Furthermore, it provides links to national high-performance computing and large-scale data analysis infrastructures, e.g. the HLRN federation [44]. GWDG will co-lead Task Area 2 and contribute to Task Areas 1 M2&4, 3 M1-3, 4 M4&5 and 5 M1.

HITS – Heidelberg Institute for Theoretical Studies gGmbH (co-applicant)

The Heidelberg Institute for Theoretical Studies was established in 2010 by SAP co-founder Klaus Tschira (1940-2015) as a private, non-profit research institute. HITS conducts basic research in the natural sciences, mathematics and computer science with a focus on the processing, structuring, and analysing of large amounts of complex data and the development of computational methods and software. The Scientific Databases and Visualization Group within HITS maintains the SABIO-RK reaction kinetics database and the FAIRDOMHub data management system and data commons. It is active in the standardisation community.

Role: HITS will contribute their experience in serving projects with tools for the first mile of data management. This includes the RightField tool for data annotation, expertise with openRefine, and fast-to-configure tabular data checkers, as well as SEEK software and the FAIRDOMHub. HITS will co-lead Task Area 4 and contribute to Task Areas 1 M1, 2 M1-3 and 3 M1&2.

IPK - Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung Gatersleben (co-applicant)

The IPK's research program and services contribute materially to conserving, exploring and exploiting crop diversity.

Its research goals are driven by the need for an efficient and sustainable supply of food, energy and raw materials, thereby addressing global ecological challenges. This involves the exploitation of the genetic diversity preserved in the Federal Central ex-situ Gene Bank and its development to a "Biological Digital Resource Centre" [45]. The research group Bioinformatics and Information Technology at the IPK is engaged in the management of research data, the implementation of integrated biological information systems/data warehouses, the development of systems for information retrieval as well as in the provision of bioinformatic tools.

Role: The IPK Gatersleben will contribute its experience in managing plant biodiversity data and plant phenotyping data as well as the collaboration within national networks like de.NBI and international networks like the European Cooperative Programme for Plant Genetic Resources (ECPGR, see letter of support 7) [11]. Together with NFDI4Earth and NFDI4Agri IPK will lead the cross-cutting use case 2 on the molecular diversity of plants and contribute to use case 10. It will provide training courses and provide access to the ELIXIR Plant community [46]. IPK co-leads Task Area 4 and contributes to Task Areas 1 M1&4, 2 M1-3, 3 M1-3 and 5 M1.

SNSB – Staatliche Naturwissenschaftliche Sammlungen Bayerns (co-applicant)

The Staatliche Naturwissenschaftliche Sammlungen Bayerns (SNSB) is one of the largest research institutions for natural history in Germany. The taxonomic facility encompasses five State Collections, the Botanical Garden Munich-Nymphenburg and eight museums with public exhibitions, located in various regions of Bavaria. The SNSB has large zoological, anthropological, paleontological and mineralogical collections and a herbarium (almost 35,000,000 specimens) as well as a biobank (DNA, tissues). The SNSB IT Center is the institutional repository for scientific data of the SNSB natural history collections [47]. The mission comprises research activities in the field of biodiversity informatics and software engineering in the field of bio- and geodiversity data management and processing.

Role: As one of the data centers of the GFBio network, the SNSB will be a data center for archiving and publishing research data within NFDI4BioDiversity. It will cooperate with scientists and organisations to mobilise high-quality data from paleontological, zoological, botanical and mycological research communities (taxonomies, collections, trait data, scientific workflow data from field and laboratory studies). It will support several national and regional volunteer expert science projects, e.g. the AraGes e.V., the AG Bayernflora, wild bees monitoring with Diversity Workbench. The SNSB will mobilise trait data from LIAS [48] and DEEMY [49]. SNSB will contribute to use case 4, 9 and 16. It will co-lead Task Area 3 and contributes to Task Areas 1 M2&3, 2 M3, 4 M1-4 and 5 M1.

UFZ - Helmholtz Centre for Environmental Research (co-applicant)

UFZ is a member of the Helmholtz association with around 1200 employees. It is constituted of 36 departments covering all aspects of environmental research of which seven are dealing with biodiversity-related topics. The UFZ has a centralized IT-department with a strong connection to the research data management (RDM) team supplying central IT- and RDM services. UFZ coordinates the eLTER RI ESFRI [50], the Helmholtz (HGF) Initiatives Climate Adaption and Mitigation (HI-CAM) and Modular Observation Solutions for Earth Systems (MOSES) [51]. UFZ contributes to TERENO [52], Digital Earth [53], ESM [54] and is a long-term contributor to ICOS [55] and LTER-D [55] (see letters of support 1 & 8).

Role: UFZ will run the eLTER use case (1) by providing the interfaces to the European scale, it will provide the ModWebBiodiv platform, connect NFDI4BioDiversity with the HGF and LTER communities for awareness building, motivation and feedback. UFZ will further lead the use case 20 (research server) and contribute to use case 11 and 12. It will co-lead Task Area 1 and contribute to Task Areas 2 M1&2, 3 M3 and 5 M1.

UniGießen - Justus-Liebig-Universität Gießen (co-applicant)

The Justus Liebig University Gießen was founded in 1607 and is the second-largest university in Hessia. Alexander Goesmann holds the chair for Systems Biology and is the coordinator of the BiGi Service Center for Microbial Bioinformatics and leads the special interest group on ELIXIR Cooperation within de.NBI. His group has established a comprehensive bioinformatics software platform with a focus on web-based database systems for genome and post-genome research. Research of Birgit Gemeinholzer focusses on plant evolution, population genomics, metabarcoding, and data management. She is specialized in the field of population genomics, technology development for accelerated biodiversity assessment and storage, and resort-specific nature conservation actions.

Role: UniGießen offers access to their IT infrastructure and software applications to support selected use cases and the development of NFDI-RDC. This will include dedicated training courses and further educational measures. UniGießen provides access to a central registry of standardized software containers and bioinformatics workflows that can be executed based on our Kubernetes infrastructure. UniGießen will also contribute to governance and monitoring tasks related to the ongoing development of the European Open Science Cloud. The Hessian research data infrastructures (HeFDI) which anchors research data management at the participating universities will support NFDI4BioDiversity. This includes technology as well as advice and services. UniGießen will contribute to use case 7, 8, 13 and 14. It will contribute to Task Areas 2 M2, 3 M1&2 and 5 M1 with Birgit Gemeinholzer co-leading Task Area 1 and Alexander Goesmann co-leading Task Area 4.

UniJena/iDiv - Friedrich Schiller University Jena (co-applicant)

The Friedrich Schiller University Jena is the largest university in Thuringia. The Michael-Stifel-Center (MSCJ) is a cross-cutting structure fostering data-intensive science in all disciplines. Biodiversity is one of the research foci at UniJena. It is one of the three universities which form together with the UFZ the core of the DFG Research Center for Integrative Biodiversity Research, iDiv. With several hundred scientists iDiv is a hotspot of biodiversity research in Germany. At UniJena two co-spokespersons are involved: Birgitta König-Ries is managing director of MSCJ and provides data management of large biodiversity projects including the Biodiversity Exploratories, the Jena Experiment and as well as the development of the BEXIS2 data management platform. Aletta Bonn, head of ecosystem services, holds a joint professorship of the Friedrich Schiller University Jena and the UFZ as part of iDiv. Her research focuses on integrative methods of biodiversity change and ecosystem service analysis at the interface of science, society and policy. She has led/co-authored several ecosystem assessments with IUCN UK [56], TEEB [57] and IPBES [58] and developed the Greenpaper Citizen Science Strategy 2020 for Germany and the Living Atlas - Nature Germany scoping study. Currently, she leads the "Analysing trends in German species data" sMon thematic network [59] and has a strong standing in the citizen science community as well as with agencies and natural history societies.

Role: Birgitta König-Ries will contribute to the use cases 22 and 23. These use cases will contribute to the NFDI-RDC in Task Area 4, where she will also be involved in leveraging her expertise on LOD. She and Aletta Bonn will be the main implementation partner for the ALA/LAND use case 23 together with BGBM. This will be part of the Jena Data Center, which will also host instances of the BEXIS2 data management platform. In this context, she will contribute to early data mobilisation in Task Area 1 M1-3 and networking in Task Area 2 M2. Finally, she will lead Task Area 3 and contribute to Task Areas 4, M2 and 5 M1. Aletta Bonn will be our main contact point in coordinating the authorities, associations and citizen scientists/volunteers and the respective use cases 9, 11, 12, 15, 17, 18, 19, 21 and 23. She will lead Task Area 1 and contribute to Task Areas 2 M2, 4 M1 and 5 M1.

UniMarburg - Philipps-Universität Marburg (co-applicant)

The Philipps-Universität Marburg offers a unique interdisciplinary biodiversity research environment comprising more than 17 working groups which focus on ecology, climate research and computer science. Over the last decade, research has focused on the development and adaptation of data-driven approaches to address core research questions in biodiversity. The database research group of the Philipps-Universität is heading the development of the VAT system that serves for the visualization, analysis and transformation of biodiversity and environmental data. In addition, UniMarburg has founded a local service center for e-research to support its scientists in all aspects of digitally assisted research.

It has the lead for the joint hessian-wide project "Hessian Research Data Infrastructure" (HeFDI) [60], which fosters coordinated research data support across 11 universities and universities of applied sciences. UniMarburg is also one of 15 universities within the Data Literacy Education network driven by the Stifterverband (see letter of support 12).

Role: Within NFDI4BioDiversity, UniMarburg will develop a seamless integration of VAT and Jupyter Notebooks. We expect that this unique combination of Python programming and a powerful visualization platform will be very appealing to (data) scientists. The database group is also responsible for the development of the container layer offered within NFDI4BioDiversity. The database group is dedicated to the implementation of metrics for measuring the involvement of researchers in NFDI4BioDiversity and other NFDIs. These metrics will serve as incentives for researchers that are involved in improving research infrastructures. The database group will lead Task Area 4 and contribute to Task Areas 1 M4&5, 2 M2 and 5 M1. In Task Area 1, we provide sources to serve the use cases with training and education. We will integrate and publish training material for case studies in the training and data literacy platform and foster the training case studies at universities. Also, we will network with other consortia on this topic.

Table 2.2.1: Summary of NFDI4BioDiversity's participants, their roles and contributions

Abbreviation	Institution	Role in NFDI4BioDiversity	Task Areas	Measures	Year
AraGes-Spinnen	Arachnologischen Gesellschaft e.V.	AraGes-Spinnen is a natural history society (spiders) leading use case 16. It connects the society including monitoring and collection data about spiders with NFDI4BioDiversity.	TA 1 TA 3 TA 4	M1 M2 M3 M2 M3 M2	2-3
AWI	Alfred Wegener Institute - Helmholtz Center for Polar- and Marine Research Bremerhaven	AWI will be a strong partner for the exchange of data and data products with NFDI4Earth. It will contribute to international standardisation committees as well as ontology development. It will link out to EBV, NEON and ESIP. It contributes to **use case 5.**	TA 2 TA 4	M1 M3 M2 M4	2-5
BGBM-DI	Freie Universität Berlin, Botanic Garden and Botanical Museum Berlin, Research Group Diatoms	BGBM-DI leads **use case 14** to connect metabarcoding data to existing curated as well evaluated research and collection data. It contributes to **use case 7**.	TA 1 TA 2 TA 3 TA 4	M1 M2 M2 M2 M2	3-5
BIOfid	FID Biodiversitätsforschung an der Goethe Universität Frankfurt	BIOfid is an expert in text-mining of biodiversity literature. It will contribute by mobilising and structuring of data from books, journals and papers.	TA 1 TA 2 TA 3	M2 M1 M3	3-5
BSH	Bundesamt für Seeschiffahrt 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.	TA 1 TA 4	M1 M1 M2 M4	3-5
BUND	Bund für Umwelt und Naturschutz Deutschland e.V. – Friends of the Earth Germany	BUND is a large environmental association supported by about 620,000 people. BUND will serve as an interface to citizen science volunteers and their needs.	TA 1 TA 2	M1 M4 M2	1-5
DDA-Vögel	Dachverband Deutscher Avifaunisten e.V.	DDA is the largest and oldest ornithological umbrella organization in Germany, coordinating the nationwide collection of data on bird diversity, abundance and distribution. It provides a data integration examples, expertise on training and outreach, user feedback (managing large communities).	TA 1 TA 2 TA 4	M2 M4 M5 M2 M2	1-5
DNAquaNet	DNAquaNet, Aquatische Ökosystemforschung (COST Action), Universität Duisburg-Essen	DNAquaNet consists of about 600 participants from 50 countries aiming to advance DNA-based tools for bioassessment and monitoring of aquatic ecosystems for legally binding biodiversity assessment at national and international level. It leads **use case 7** and will provide large genetic and genomic data sets for biodiversity assessments of aquatic ecosystems.	TA 1 TA 2 TA 3	M1 M2 M2	3-5

DFG

Abbreviation	Institution	Role in NFDI4BioDiversity	Task Areas	Measures	Year
GdO-Libellen	Gesellschaft deutschsprachiger Odonatologen e.V.	GdO-Libellen is a natural history society (dragon flies) providing **use case 15**. It connects the society and their data about dragonflies with NFDI4BioDiversity.	TA 1 TA 2 TA 3 TA 4	M1 M4 M5 M1 M2 M2	2-4
GfI-Fische	Gesellschaft für Ichthyologie e.V.	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.	TA 1 TA 4	M1 M2	3-5
GfÖ	Gesellschaft für Ökologie e.V.	GfÖ is one of the world's largest societies for scientific ecology and opens access to a large community. It will support training, outreach, user feedback and cultural change.	TA 1	M4 M5	1-5
HIFMB	Helmholtz Institute for Functional Marine Biodiversity at the University of Oldenburg	HIFMB has a focus on marine biodiversity change and marine conservation. It leads **use case 5** as well as tools for data analysis and data integration, addressing use cases and long-term data analysis. Provision of data and expertise in data science.	TA 1 TA 3 TA 4	M2 M3 M2	2-5
HLNUG	Hessisches Landesamt für Natur, Umwelt und Geologie	HLNUG is a federal state environment agency leading **use case 13**. Provision of biodiversity data stored in MultiBase CS information system. It contributes to networking of stakeholders (including authorities, citizen science) in nature conservation.	TA 1 TA 4	M1 M2	3-5
IfL	Leibniz Institut für Länderkunde	IFL is a research institute for regional geography leading **use case 19**. IfL will develop visualizations of various datasets and test the usability with user groups.	TA 1 TA 4	M1 M2	2-3
IGB	Leibniz-Institut für Gewässerökologie und Binnenfischerei	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 of comprehensive institutional data management practices, and in establishing an state of the art infrastructure.	TA 1 TA 2 TA 4	M2 M1 M2	3-4
InfAI	Institut für Angewandte Informatik	InfAI provides terminology services for the NFDI-RDC measure on data integration and harmonization.	TA 4	M2	1-5
IÖR	Leibniz-Institut für ökologische Raumentwicklung	Leibniz Institute of Ecological Urban and Regional Development (IÖR) will describe and analyse land use and land cover as one import factor moulding biodiversity. IÖR will lead **use case 6**.	TA 1 TA 2 TA 4	M1 M1 M1	3-5

DFG

Abbreviation	Institution	Role in NFDI4BioDiversity	Task Areas	Measures	Year
JKI	Julius Kühn-Institut Bundesforschungsinstitut für Kulturpflanzen	JKI provides data from long-term research in the fields of soil science, plant cultivation, genetics, plant breeding and plant protection. German long-term and permanent field trials on JKI experimental stations, demonstration areas, on-farm investigations of cultivated plants, pathogens, useful insects or indifferent species.	TA 2 TA 4	M1 M1 M2	3-5
LAU	Landesamt für Umweltschutz Sachsen-Anhalt	LAU is a federal state environment agency leading **use case 11**. It will provision biodiversity data stored in the MultiBase CS information system. It contributes networking of stakeholders (including authorities, citizen science) in nature conservation.	TA 1 TA 4	M1 M2	3-5
LfULG	Sächsisches Landesamt für Umwelt, Landwirtschaft und Geologie	LfULG is a federal state environment agency leading **use case 21**. Provision of biodiversity data stored in MultiBase CS information system. Contributes networking of stakeholders (including authorities, citizen science) in nature conservation.	TA 1 TA 4	M1 M2	2-4
MfN	Museum für Naturkunde Berlin	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.	TA 1 TA 2 TA 3	M3 M1 M3	1-5
MPI-Biogeochem	Max-Planck-Institut für Biogeochemie	MPI-Biogeochem will provide a strong link to NFDI4Earth and will lead the iDiv **use case 22** showing the benefits of close integration of data across domains.	TA 1 TA 2 TA 3 TA 4	M1 M1 M3 M2	2-4
Naturgucker	Naturgucker gemeinnütziger e.G.	naturgucker.de leads **use case 8** and is well established within citizen sciences scene in Germany (~70,000 registered users, more than 200,000 webusers p.a.). They contribute experience in handling and evaluating observation data. naturgucker has one of the largest collections of nature observation data over all species groups.	TA 1 TA 2 TA 4	M2 M3 M4 M5 M2 M2	3-5
NetPhyD-Pflanzen	Netzwerk Phytodiversität Deutschland	NetPhyD-Pflanzen is the largest plant natural history society and will lead **use case 17.** It connects the society including monitoring and collection data about plants with NFDI4BioDiversity.	TA 1 TA 4	M1 M2	2-4

DFG

Abbreviation	Institution	Role in NFDI4BioDiversity	Task Areas	Measures	Year
NP Bayerischer Wald	Nationalparkverwaltung Bayerischer Wald (Behörde des Bayerischen Staatsministeriums für Umwelt und Verbraucherschutz)	NP Bayerischer Wald is a national parc authority leading **use case 12**. NP Bayerischer Wald is representative for other NP's generating different types of long-term monitoring data without an own information system.	TA 1 TA 4	M1 M1 M2	3-5
NP Hunsrück-Hochwald	Nationalparkamt Hunsrück-Hochwald	NP Hunsrück-Hochwald is a national park authority leading **use case 20**. NP Hunsrück-Hochwald represents three NP's running the research server serving as a meta data information system which will be connected to NFDI4BioDiversity.	TA 1 TA 4	M1 M2	3-5
SAB	Staatliche Archive Bayerns	SAB 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.	TA 2 TA 3 TA 4	M1 M2 M3 M2	3-5
SGN	Senckenberg Gesellschaft für Naturforschung	Natural history collection and GFBio data center for archiving and publishing biodiversity data. SGN contribute curatorial and technical services and expertise. SGN leads **use cases 9** for ecological analyses, contributes to **use case 1** and cooperates with other NFDIs and international political networking activities.	TA 1 TA 2 TA 3	M1 M2 M3 M1 M2 M1 M3	1-5
SMNS	Staatliches Museum für Naturkunde Stuttgart	Natural history collection and GFBio data center for archiving and publishing biodiversity data. SMNK 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 partner in the AraGes **use case 16**.	TA 1 TA 3	M1 M3 M3	1-5
SUB	Niedersächsische Staats- und Universitätsbibliothek Göttingen	SUB is one of the largest libraries in Germany, a leader in the development of digital libraries and involved in national and international networks and other NFDIs. SUB contributes to research data management by training measures, community specific workshops and training materials.	TA 1 TA 2 TA 4	M3 M4 M1 M2 M1 M2 M4	1-5
Thünen-GDI	Thünen Institute	Thünen-GDI leads **use case 10** which connects NFDI4Agri and NFDI4BioDiversity by providing the biodiversity monitoring as a testbed for data-integration.	TA 1 TA 2	M1 M1 M2	3-5
UniGöttingen	Universität Göttingen	UniGöttingen represented by the "Daten Lesen Lernen"-Project, has the role of a user- and community-engagement facilitator. By using the infrastructures and data that will be provided by NFDI4BioDiversity and developing of resultant case studies in the data literacy education program.	TA 1	M4 M5	1-5

DFG

Abbreviation	Institution	Role in NFDI4BioDiversity	Task Areas	Measures	Year
UniLeipzig	Universität Leipzig	UL will contribute to the iDiv **use case 22** integrating a wealth of plant related data hosted at iDiv with environmental data provided by NFDI4Earth.	TA 1 TA 2 TA 3 TA 4	M1 M1 M3 M2	1-4
ZFMK	Zoologisches Forschungsmuseum 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. ZFMK is 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.	TA 1 TA 3	M1 M3 M3	1-5
ZMT	Leibniz Zentrum für Marine Tropenforschung	ZMT will provide unique long-term time series from tropical regions. They will contribute to the research data management infrastructure by providing curation capacities.	TA 1 TA 4	M2 M3 M2	3-5

DFG

How will the network gain the expertise it needs to implement its work programme if this expertise is not available within the consortium?

For year 4 and 5, a total of 500,000 Euro from the centrally managed funds have been set aside to be able to welcome additional participants at the lifetime of the consortium. We expect that NFDI-wide rules will be worked out for the evaluation and admission of additional participants as described in detail in section 2.5.

2.3   The consortium within the NFDI

NFDI4BioDiversity contributions to NFDI

By addressing the biodiversity communities, NFDI4BioDiversity is unique in the ecosystem of emerging NFDI consortia. With our particular set of universities, research institutes, collections and repositories as well as learned societies, associations, authorities engaged in nature protection from parks to spiders, beetles and dragonflies, NFDI4BioDiversity will contribute its view on the diversity of organismic life, interactions of organisms as well as between organisms and the environment to NFDI (Figure 2.3.1). NFDI4BioDiversity is part of the NFDI4Life-Umbrella, where the main roadmap in addressing the big challenges of a collaborative cross-domain infrastructure for research data management has been worked out. We have and will go on to coordinate our approaches and strategies with other consortia by initiating and participating in meetings and bilateral communications for a productive exchange within NFDI. As a matter of fact, NFDI4BioDiversity has initiated the "Berlin Declaration on NFDI Cross-Cutting Topics" and started the discussion with other consortia on the exchange of data, data products, services and expertise between related domains. This said, all consortia involved in the Berlin Declaration have unequivocally expressed their willingness to continue the process, but also stated that around two to three "NFDI" years are needed until the details on data and service exchange between consortia are flashed out. The following list of coordinated actions should therefore be taken as examples and not as a comprehensive list.

Coordinated actions with other consortia

National Research Data Infrastructure for Earth System Science. NFDI4Earth is a natural partner in terms of common data and metadata models and related geo-standards, the re-use of infrastructure services, i.e. in data mobilization and cloud systems. It is a consumer of domain e.g. specific (taxonomic) terminology services like names of fossils and stratigraphic concepts and will be a partner for the shared data repository PANGAEA. Already in the application phase, the speakers worked closely together to shape complementary applications and to establish a common understanding about governance and finances.

We host 10 members that are shared between the consortia, which will be anchors for an ongoing mutual understanding when working together to create ONE NFDI. Data exchange between NFDI4Earth and NFDI4BioDiversity is essential to better understand the impact of climate change on biodiversity loss and vice versa.

National Research Data Infrastructure for Agricultural Sciences. With NFDI4Agri there are considerable overlaps in investigated objects e.g. plants, fungi or animals. Information about these organisms is stored in repositories, which should be made accessible and usable by NFDI4BioDiversity as well as NFDI4Agri. The consortia differ in terms of their user community. In NFDI4Agri the data consumers are for example agricultural and soil scientists. However, there are significant synergy effects regarding the development and harmonisation of access opportunities. Specifically, collaborations in the areas of object identification, phenotypic trait collection and genotyping data management are feasible. Again, a good set of shared members in both consortia guarantee an efficient flow of information and mutual interactions.

National Research Data Infrastructure for Personal Health Data. Common interests with NFDI4Health are biobanking and sample management as well as common data types like molecular "Omics" data. The legal framework and FAIR metrics that will be developed in the health and medical field in compliance with privacy regulations and ethics principles are of interest for the subset of sensitive data provided by NFDI4BioDiversity e.g. on endangered species. Similar to NFDI4Earth, the deeper understanding of wellbeing, resilience and health of each of us is closely linked to the stability of the ecosystem, where biodiversity is a major factor.

NFDI Consortium of the Medical Informatics Initiative (MII) and the German Centers for Health Research (DZG). The NFDI4BioDiversity and NFDI4Medicine consortia are located in neighbouring scientific domains with a common interest in linking biodiversity, health and medical research data. Research data on microbial diversity (e.g. the gut and skin microbiome) becomes increasingly relevant for medical research and personal health. Sharing of biodiversity, environmental (NFDI4Earth) and health data together with data from other domains such as animal health and agricultural sciences (NFDI4Agri) should, in the long run, lead to a broad ONE health approach across the disciplines. Concrete collaboration is therefore planned on the topics of metadata harmonization and interoperability across domains.

National Research Data Infrastructure for Chemistry: Out of the data managed by NFDI4Chem secondary metabolites and metabolomics data is of particular interest for the biodiversity community. Integrated data access across the NFDIs is thus crucial. In addition, collaboration is planned on the evaluation and development of data management tools for the early stages of the data life-cycle.

Collaboration with **KonsortSWD** is already projected to exchange regional economic data to contextualise biodiversity studies as well as knowledge on anonymisation.

NFDI4RSE which is scheduled for submission in a later round, aims for sustainable management of research software in the context of research data management, a topic with clear links to our Task Area 3.

Members of NFDI4BioDiversity participating in other consortia
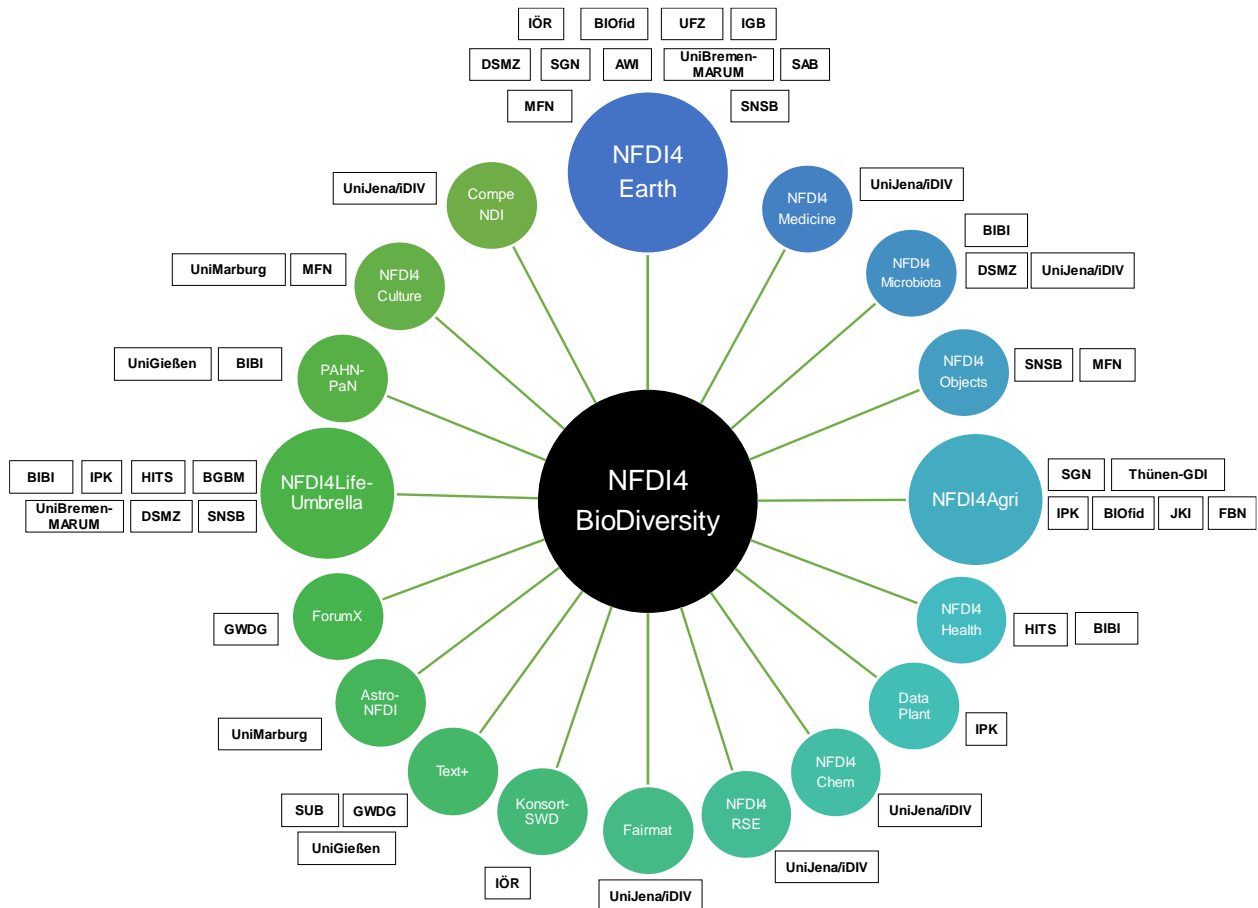
Figure 2.3.1: NFDI4BioDiversity partners and their interactions with other NFDIs

Expectations with regard to coordination of topics within NFDI

Berlin Declaration on NFDI Cross-Cutting Topics

A set of cross-cutting topics has been agreed and signed by 21 NFDIs which need to be coordinated within NFDI. They have been filed as the "Berlin Declaration on NFDI Cross-Cutting Topics" [22].

From the NFDI4BioDiversity perspective the following topics have been defined as most urgent and relevant to be addressed and decided consortia-wide within the NFDI's collaborative framework:

(Meta)data harmonisation and interoperability across domains

Easy exchange of (meta)data across NFDIs/domains would leverage new research potential e.g. by combining biodiversity, environmental, and even social science and health data. We will address structural, semantic, and conceptual hurdles for the harmonization of data and metadata.

The overall goal is to minimize the necessary efforts. For this purpose, we will concentrate on standards champions in the various domains and work towards common standards. We will follow the schema.org [61] principles by building on a common core schema with community extensions like bioschemas.org. For semantic (meta)data harmonization we will need terminology services as described below. This cross-cutting topic will be addressed in Task Areas 2 and 4.

Terminology management and services

Terminology services provide the basis for semantically enriched data management from retrieval to archiving by integrating and harmonizing heterogeneous terminological resources (incl. taxonomies). In cooperation with all NFDIs, we want to extend our terminology service [62], developed for the biological and environmental domain, to provide services and tools to find, explore, share and reuse terminologies for the semantic enhancement and harmonization of data across domains. A major point of action that can only be solved cooperatively, will be the mapping between terminologies from different domains. This cross-cutting topic will be addressed in Task Areas 2, 3 and 4.

Research data commons

NFDI4BioDiversity is planning the implementation of the NFDI Research Data Commons to pool data and applications to leverage data science for the different stakeholders. This NFDI-RDC is conceived as a virtual expandable infrastructure that allows users to store, analyse, share data and results and to combine diverse data types. Together with other NFDIs, we will explore if NFDI-RDC can be extended and serve as a common platform for data sharing and cross-domain data analysis. (Meta)data harmonization as well as terminology management, as described above, are building bricks for the NFDI-RDC. A particular problem to be addressed when aggregating data from different sources is provenance. This cross-cutting topic will be tackled in Task Area 2 and 4.

Certification systems and strategies

To meet the increasing demands of funders, publishers, and research organisations to get the quality of data and services formally accredited, certifications by organisations like the ICSU-WDS [63], DINI e.V. [64], nestor [65], or the new CoreTrustSeal [66] have become widespread means. Based on our long-term experiences with the certification of the PANGAEA information system [35], our essential role in the development of the CoreTrustSeal, and our involvement in the H2020 FAIRsFAIR project [67], NFDI4BioDiversity is in a prime position to share this knowledge, develop strategies for certification and guide the certification process in other NFDI consortia as well. Certification will address authenticity, integrity, confidentiality, and availability of data and services as well as the assessment of the FAIRness of data centers and their holdings. This cross-cutting topic will be addressed in Task Areas 2 and 3.

Graduate Education

NFDI4Health, NFDI4BioDiversity, NFDI4Earth together with the Federal State of Bremen and the University Bremen Research Alliance (UBRA) [68], will establish a cross-domain graduate education programme on research data management and data science.

This programme starting in November 2019 at the University of Bremen, will serve as a blueprint for the NFDI in general. The curriculum and modules developed, pre-tested and refined according to the feedback of the students will be subsequently provided to all NFDI consortia. The material will be tailored to the needs of graduate students working on the research fields addressed by the NFDIs in year one. It will be evaluated by the consortia and our user communities in year two. In year three, the material will be revised and extended. The NFDI-wide refinement and roll-out is planned for year four and five, respectively. This cross-cutting topic will be addressed in Task Areas 1 and 2.

Governance & sustainability

Suitable governance structures are key to ensure sustainable operations of a distributed infrastructure like NFDI. Therefore, a major challenge for NFDI and/or the NFDI consortia and directorate will be the identification of an appropriate legal entity which serves the interests of the consortia and service providing host institutions. With the foundation of the GFBio association (e.V.) as a not-for-profit legal entity in 2016, NFDI4BioDiversity has ample experience in exploring possible legal forms as well as business models. We would be happy to share this knowledge with all NFDI consortia, the DFG and the Directorate to create a common understanding of the pitfalls and challenges to work towards a common model for governance and sustainability. Together with NFDI4Earth, we are currently discussing a cross-NFDI 'collaborative governance' model. This cross-cutting topic will be addressed in Task Areas 2 and 5.

2.4   International networking

Biodiversity research and monitoring is a global concern, with conservation priorities, biological threats, and opportunities to access natural capital knowing no borders. Data and information must flow accordingly: regardless of local success, impacts will be severely limited without well-developed interfaces to coordinated international efforts. Thus, participation and leadership in international networks with continuous strategic alignment will be a cornerstone activity in NFDI4BioDiversity. This engagement extends beyond biodiversity-specific networks, also targeting networks in Earth sensing and socioeconomics, which both impact and are impacted by biodiversity. Dedicated and generic networks handling biodiversity data are many. To prevent spreading resources too thinly, NFDI4BioDiversity's networking activity will promote greater cross-network coherency rather than simply maximizing involvement. Below, we briefly describe selected networks within which we see NFDI4BioDiversity establishing itself as a sustained stakeholder and contributor.

Networking with international infrastructures and initiatives is implemented in Task Area 2, which ensures that NFDI4BioDiversity interacts properly with the relevant actors, resulting in a coordinated service infrastructure and exploiting the possibilities of mutual data use.

Biodiversity-centric networks

The Group on Earth Observation's Biodiversity Observation Network (GEO BON) [6]: GEO BON has shown stability and structured growth along both regional (e.g. the Arctic BON) and thematic (e.g. the Marine BON) dimensions. Further, GEO BON working groups - with leadership from Germany's iDiv centre (represented by UFZ, DSMZ, IPK and UniJena/iDiv) - have engaged in a mission to formulate a set of Essential Biodiversity Variables (EBVs) [69] as a complement to Essential Variables in climate (ECVs), oceans (EOVs), and weather (EVs). NFDI4BioDiversity offers Germany the opportunity to interface with GEO BON at a national scale, and to be a key contributor to a regional European BON. With this model, consolidated efforts to test, co-develop, and build capacity on delivering the EBVs can be realised, as well as structured links to biodiversity policy activities.

The Long-term Ecological Research Sites networks: The US LTER network [8] and its international (iLTER) [70] and European counterparts (eLTER) [9] are key points of interaction for NFDI4BioDiversity (see letter of support 8). Their persistence has led to impressive field-tested digital infrastructures and policies in a multi-partner management context. Further, they - and associated initiatives such as ALTER NET [71] - serve to inform multiple levels of the policy domain, which can be leveraged in the German national context. NFDI4BioDiversity co-applicants' (e.g. UFZ) existing links and contributions to eLTER will nucleate an infrastructure-to-infrastructure interface. The Long-term Ecological Research Sites networks: The US LTER network [8] and its international (iLTER) [70] and European counterparts (eLTER) [9] are key points of interaction for NFDI4BioDiversity. Their persistence has led to impressive field-tested digital infrastructures and policies in a multi-partner management context. Further, they - and associated initiatives such as ALTER NET [71] - serve to inform multiple levels of the policy domain, which can be leveraged in the German national context. NFDI4BioDiversity co-applicants (e.g. UFZ) existing links and contributions to eLTER will nucleate an infrastructure-to-infrastructure interface.

The National Ecological Observatory Network (NEON) [10]: As a multi-decade and continental-scale project, NEON is developing new norms and best practices at the intersection of research-focused and operational ecological observation. It monitors the impacts of land-use change and climate variation on the biosphere, as well as their links to the geosphere, hydrosphere, and atmosphere. Its ability to integrate a wide range of advanced and traditional methods and technologies, alongside their diverse data flows in a production-grade context is of high interest to the mission of NFDI4BioDiversity.

NFDI4BioDiversity proposers like AWI have links to NEON through networks such as ESIP (see below), which will be nurtured to share capacities focused on production-grade infrastructure management (see letter of support 1).

The European life-sciences Infrastructure for biological Information (ELIXIR) [16]: ELIXIR is an intergovernmental organisation that brings together life science e-resources from across Europe. ELIXIR coordinates 22 members, bringing together over 220 research organisations. NFDI4BioDiversity co-applicants BIBI, UniBremen-MARUM, HITS, DSMZ and IPK have well-established links to ELIXIR, particularly in the molecular life sciences, contributing core resources and pioneering new links between previously siloed sub-domains of biodiversity science such as DNA-based community ecology and global taxonomic data resources [72].

The Consortium of European Taxonomic Facilities (CETAF) [4]: CETAF is the world's largest association of natural history organisations (botanical gardens, natural history museums, natural science museums, seed banks, etc.). The 61 member organisations together hold 1.5 billion collection objects, covering 80% of the species described. Several members of the NFDI4BioDiversity Consortium are actively involved in CETAF (as institutional members and in the working groups, for example as Chair of the CETAF Information Science and Technology Committee and as members of the Executive Committee). CETAF can thus be seen as an important strategic partner for networking with the international collection community (see letter of support 3).

The Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) [58], assesses the state of biodiversity and the ecosystem services it provides to society, in response to requests from decision-makers. The IPBES Plenary established a task force on knowledge and data (decision IPBES-2/5, section II) with the primary purpose to support the implementation of deliverables 1(d) and 4(b): Priority knowledge and data needs for policymaking addressed through catalysing efforts to generate new knowledge and networking and the delivery of an information and data management plan. NFDI4BioDiversity will seek to interface with such task forces to better integrate national biodiversity data products into the IPBES process.

Generic networks

Echoing our natural link to NFDIs in Earth and societal observation, our international networking activities must also position NFDI4BioDiversity within broader networks to identify, curate, shape, and share emerging capacities for our stakeholders' benefit. Below, we list a selection of generic networks we have identified as key targets for our networking strategy.

European Open Science Cloud (EOSC) [21]: Although EOSC is a European infrastructure, it is designed to fit and serve the global landscape. EOSC is meant to link a multitude of existing and future research data infrastructures to provide added value services for data providers and users.

NFDI4BioDiversity and its cloud-based services will be designed to be compatible with the evolving EOSC services and thus provide the basis for contributions to EOSC and use of the relevant EOSC services. Furthermore, NFDI4BioDiversity offers an ideal basis for the calibration of EOSC services with its broad content orientation and the already large volumes of existing research data.

The Research Data Alliance (RDA) [12]: RDA was launched as a community-driven initiative in 2013 by the European Commission, the United States Government's National Science Foundation and National Institute of Standards and Technology, and the Australian Government's Department of Innovation with the goal of building the social and technical infrastructure to enable open sharing and re-use of data. In recent years, the RDA has grown to become one of the foremost fora to coordinate global data science across disciplines and stakeholders, featuring over 8800 members from 137 countries. Several of our partners are chairing and contributing to RDA interest and working groups and, collectively, NFDI4BioDiversity can help establish a much-needed disciplinary alliance around biodiversity [73]. This will provide an excellent forum to establish national scale working groups on biodiversity, to better align our infrastructure with those around the globe.

Data Observation Network for Earth (DataOne) [19]: DataOne is an NSF funded distributed network for environmental and earth observation data. Like NFDI4Biodiversity, DataOne offers a broad portfolio of software solutions and services for e.g. cataloguing and accessing distributed data sources, interoperability of environmental research data, and analysis. A special focus is the integration of the research community. NFDI4Biodiversity will ensure through its activities that services are developed and offered in a complementary and cooperative way. Equally important will be reciprocal free access to data. Joint compatibility with standards will be ensured by integrating NFDI4Biodiversity and DataOne into the relevant standardisation initiatives such as TDWG Biodiversity Information Standards (see letters of support 4 & 13) [7].

The Earth Science Information Partners (ESIP) Federation [74]: ESIP is a 501(c)3 networking body founded by NASA [75], NOAA [76], and USGS [77]. The ESIP Federation has become an international, multidisciplinary network of governmental, public, and private stakeholders working in planetary observation.

ESIP's focus on data and information products and the far-reaching influence of its participants allows to effect changes of global significance in systems such as remote sensing, geosemantics, drone standards, and data life-cycle management. Recent ESIP meetings have featured several projects handling biodiversity data. NFDI4BioDiversity partners are already active in ESIP.

DFG

NFDI4BioDiversity and other NFDIs will greatly benefit from interactions with the ESIP community, interacting with other federal-level data and information bodies to share and co-develop capacities. Environmental Research Infrastructures (ENVRI) [17]: Evolving from the original ENVRI project, through ENVRI+ and now ENVRI-FAIR, this multinational and multidisciplinary network links infrastructures across Earth observation, including the biosphere. Its objective to streamline infrastructure activities through improved coordination can be an asset to NFDI4BioDiversity, especially in conjunction with activities in ESIP and RDA.

2.5   Organisational structure and viability

Internal organisational structure of NFDI4BioDiversity

Applicant: UniBremen-MARUM is the designated applicant for NFDI4BioDiversity, providing the Speaker. The Speaker will coordinate the consortium based on due diligence by making use of the instruments, task areas and measures described below. As we see the speaker's role as overarching, UniBremen-MARUM does not lead a specific task area but will be active in the Special Interest Groups 4 and 5 (see below).

Co-applicants: Co-applicants are members of the NFDI4BioDiversity consortium named in the application. They provide the co-spokespersons. Their duties and obligations are defined in this application and kept in the Consortium Agreement (CA) which will be filed as soon as the funding has been granted. To fulfil their duties, they have an assigned budget for coordinating their task areas, as well as their specific contributions.

GFBio e.V.: The Speaker together with GFBio e.V.'s executive secretary and coordinators will govern NFDI4BioDiversity and represent NFDI4BioDiversity internally and externally by acting as the prime contact point for all data providers, users, co-applicants and participants as well as the DFG, Board of Governors, NFDI Directorate and Senate (Figure 2.5.3).

Participants: Participants are legal entities like universities, institutes, organizations, societies, associations and authorities that represent a specific user or provider community. Based on an application, participants can receive a dynamic compensation depending on their participation from the centrally managed funds hosted by UniBremen-MARUM. An initial set of 35 participants has been identified based on their expertise and contributions to NFDI4BioDiversity and are already listed in this application. Since they have been part of the review process the compensation of their contributions will be initiated as soon as the CA has been filed, recalling that most of them will not start before year two. Further participants can be added to the consortium to fulfil duties that are emerging in the lifetime of NFDI4BioDiversity. For them, an evaluation procedure will be set up and applied (see SIG4).

Task Areas Leaders and Measure Leaders: The Task Areas Leaders (TAL) are the co-spokespersons of the consortium who are responsible for coordinating and managing the work within their Task Areas (TA).

They supervise all the work done in the measures of their TA and have direct day-to-day contact with the Measure Leaders (ML). They monitor the overall workflow within their TA and take care that milestones assigned to their TA are reached in time. They are part of the decision-making in NFDI4BioDiversity through their membership in the Steering Committee (SC, see below).

Measure Leaders (ML) are co-applicants or participants of the consortium who are responsible for the implementation of the work assigned to their measures. They take care that milestones assigned to their measures are reached in time. ML shall report progress within dedicated milestones to the steering committee as well as on a day-to-day basis to the TALs and make every possible effort to solve problems occurring at the operational level.

Special Interest Groups: Special Interest Groups (SIGs) are responsible for cross-cutting topics within the consortium and later NFDI. SIGs are the think-tanks of NFDI4BioDiversity which prepare decisions for the SC. They consist of experts of the consortium for a specific topic (co-applicants and participants). If necessary, SIGs can invite experts (e.g. providers and users) from outside the consortium as temporary or permanent guests. With the start of NFDI4BioDiversity, five SIGs will be established: SIG1: Public relations, SIG2: Service monitoring, SIG3: Training, SIG4: Grants and SIG5: Strategic planning. Further SIGs will be added if necessary. The concept of SIGs was adapted from the positive experiences on governing de.NBI.

SIG1, public relations: The SIG will take care of the overall public relations concept of NFDI4BioDiversity in close cooperation with Task Area 1 and 2. This includes the webpage and social media activities to reach out and brand NFDI4BioDiversity as the trusted partner for research data management in the community. It will provide guidelines for the harmonization of PR activities across the consortium and as a part of NFDI. The SIG will be chaired by the public relations coordinator.

SIG2, service monitoring and FAIRness: The SIG will establish and monitor Key Performance Indicators (KPIs) across all NFDI4BioDiversity services and archives with respect to usability and FAIRness in cooperation with Task Areas 1, 3, 4 and 5. It will establish a service index considering the usage, user satisfaction and resources used for each service and take care of monitoring the performance over time. It will be responsible for the certification process of the archives. This SIG will be chaired by the services coordinator.

SIG3, training and education: The SIG will develop an efficient and NFDI wide complementary concept for all training and education activities offered by the consortium in cooperation with Task Area 1. The SIG will establish evaluation criteria for each training entity and take care of their implementation. The SIG will align its activities with the Bremen Graduate Education for research data management and data science as well as cross-cutting training and education activities in NFDI. The SIG will be chaired by the training coordinator.

SIG4, grants: The SIG will develop workflows, application procedures and evaluation criteria for the management of participants in cooperation with Task Area 5. It will establish measures for the eligibility of additional co-applicants and participants for receiving financial compensations from the centrally managed funds. SIG4 will ensure that the evaluation process includes an independent peer-review process. Since many consortia will have centrally managed funds a general NFDI-wide evaluation approach would be most effective. The SIG will be chaired by the Speaker.

SIG5, strategic planning and sustainability: The SIG will reflect on the overall strategic development and sustainability models of NFDI4BioDiversity in cooperation with all co-applicants. It will monitor the establishment of NFDI, as well as national and international activities with respect to research data management and a special focus on international activities as outlined in section 2.4 and Task Area 2. It will align its activities with the NFDI Scientific Senate and Directorate. The SIG will be chaired by the Speaker.

Disbursement of funds: UniBremen-MARUM will administer the money transfer from the funders to the co-applicants. It will be responsible for establishing the legal framework with all NFDI4BioDiversity co-applicants by setting up a CA. If necessary, it will also take care of the legal framework and contracts with commercial service providers and companies.

University Bremen's budget includes a share of money (centrally managed funds) for the in-time support of participants essential for the instant development of NFDI4BioDiversity. Performance grants will be provided e.g. to mediate the development of standards, templates, vocabularies, ontologies and data workflows as well as the provision of services (including public relations, media and social media) and training activities to just name some examples (Figure 2.5.1).
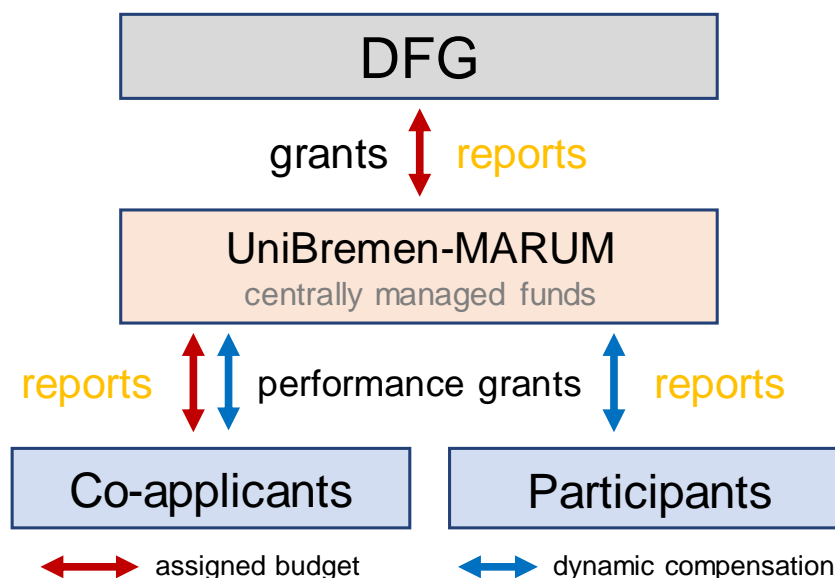

Figure 2.5.1: Funding and reporting in NFDI4BioDiversity

Financial support for co-applicants and additional participants

The application to receive funds from the centrally managed funds as a co-applicant or participant can either be initiated by the SC (e.g. suggested by the SIGs) or on request by the co-applicant or potential participant. Financial compensation will be taken from the centrally managed funds managed by the UniBremen-MARUM. To give an example: A state or federal authority or citizen science project has a set of data that needs to be mobilized and made accessible and interoperable. To compensate for these additional efforts, money can be granted. If the measures of a participant are fulfilled, they will be released. Details will be worked out by SIG4 in the runtime of NFDI.

Decision making and viability

Speaker: The Speaker manages NFDI4BioDiversity based on the decisions and suggestions of the SC, the SAB, BoG (see below) as well as the NFDI Directorate by advising GFBio e.V. To facilitate the communication the Speaker will stand for election as the Chair of GFBio e.V. As part of the consortia assembly the Speaker will place NFDI4BioDiversity related NFDI-wide topics for discussion and decision. The Speaker will support at his best the NFDI-wide cooperation and monitors that NFDI4BioDiversity meets all NFDI requirements and obligations. The Speaker represents NFDI4BioDiversity internally and externally.

Scientific Advisory Board: The Scientific Advisory Board (SAB) consists of around five scientific experts in the field of biodiversity and informatics. NFDI4Biodiversity will strive to ensure that the SAB is balanced for scientific field, gender, and nationality. Sharing of SAB members between related consortia is intended to facilitate communication and common workflows. SAB members will be proposed by the Speaker, the SC, and the GFBio e.V. members meeting. The SAB will advise on the needs and demands of the scientific community to support the further development of NFDI4BioDiversity.

Steering Committee: The Speaker, GFBio e.V. chair, GFBio e.V. executive secretary, TALs, MLs and SIG leaders (SIGL) form the steering committee (SC). The SC may decide to temporarily enlarge the SC with guests. The SC must decide unanimously on the voting rights of these additional SC members. No more than three additional SC votes are allowed. In case an SC member has more than one function he/she has only a single vote with the exception if the Speaker and the GFBio e.V. chair is the same person two votes are assigned. For votes, TALs, MLs and SIGLs may nominate a representative. The SC elects a chair who is responsible for calling the meetings and reporting the results to the Speaker, the executive board of GFBio e.V. and the SAB. The SC will meet in regular intervals: video conferences shall take place every month and physical SC meetings will be held at least three times a year (Figure 2.5.2).

The SC is responsible for the operation, progress monitoring and day to day decision-making in NFDI4BioDiversity. It reviews all the work done, the progress, status of milestones and discusses any problems in the project's activities. The members of the SC may modify the workflow for the better use of resources and for the best performance of the consortium in order to achieve the objectives and measures of the project. It decides on the establishment of SIGs for cross-cutting activities. The SC assigns tasks to the SIGs and decides on the outcome prepared by the SIGs. The SC will seek advice from the SAB to prepare their decisions and the tasks for the SIGs. The SC assigns tasks for execution by the GFBio e.V. Strategic decisions which affect the consortium as a whole cannot be decided by the SC. The SC may collect and submit such issues to the members meeting.

Every 30 months the SC will request progress reports from all co-applicants and participants as well as providers and users who received financial support, indicating possible deviations from the project's goals. The progress reports will be evaluated by the SC. Irregularities, deviations and delays will be subsequently reported to the Speaker and SAB. In case of severe delays, deviations or misconduct the SC reports to UniBremen-MARUM, BoG and NFDI Directorate. Based on their advice/decisions the next share of the budget will be distributed to the co-applicants and participants via UniBremen-MARUM. In particular, this means that redistribution of the budget will depend on the progress of the work, monitored by the fulfilment of KPIs and milestones.
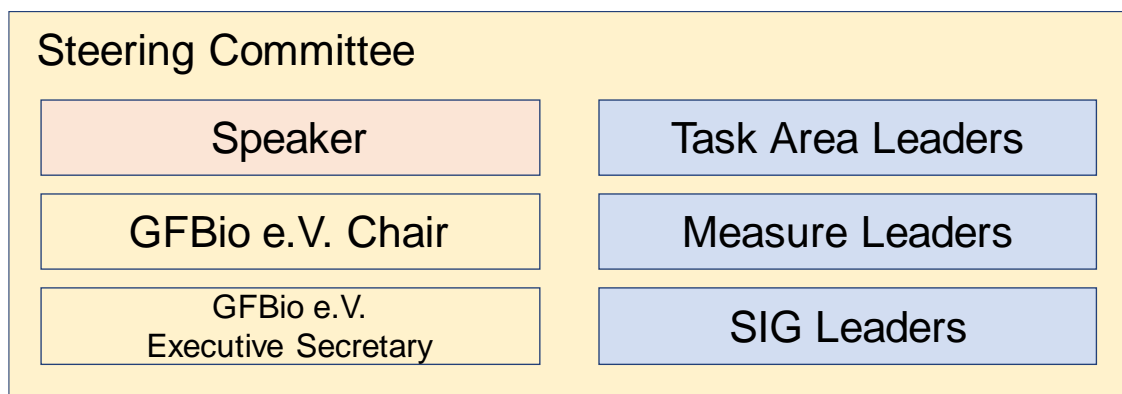
Figure 2.5.2: Composition of the Steering Committee of NFDI4BioDiversity

Board of Governors: The governance structure will be complemented by an independent Board of Governors (BoG, Aufsichtsrat) consisting of max. five persons representing different stakeholders. A possible composition could consist of one representative each of users, infrastructures, NFDI directorate, BMBF, DFG. An important task of the BoG is conflict resolution. It may propose solutions for controversial subjects or topics which affect the consortium as a whole. Most likely it will be a key component in the evaluation process for grants from the centrally managed funds (see SIG4).

BoG members will be proposed by the SC, the GFBio e.V. members meeting and the NFDI Scientific Senate and/or Directorate. The final composition of the BoG must be confirmed by the NFDI Directorate.

The BoG will be invited to attend the annual members meeting. The Speaker ensures that the BoG has full access to progress reports and project results. Sharing of BoG members among consortia or building an NFDI-wide BoG is intended.

GFBio e.V.: GFBio e.V. stands at the intersection of the Speaker, the SC, and the SAB. It is responsible for the implementation of the decisions of the SC. Together with its asset of full-time coordinators and TALs/MLs it monitors that all milestones are met. With GFBio e.V. in place, viable engagement of the co-applicants, participants, providers and users during the runtime of the consortium can be efficiently practised by the membership in GFBio e.V. With the members meeting being the highest decision body of an association, all members shape directly any decision of NFDI4BioDiversity.

The members meeting will elect the executive board (Vorstand, max. five persons), which will subsequently designate the executive secretary (Geschäftsführer). Further details on the voting rules and regulations in the member's meetings are defined in the GFBio e.V. statutes. Together with the Speaker the executive board will be the first contact point for requests from inside and outside the consortium. It will take care and track that any suggestions made by the SC are subsequently implemented by the TAs. The executive secretary together with the administration and team of coordinators are responsible for the daily operation of NFDI4BioDiversity (Figure 2.5.3).

The member's meetings will be accompanied by a meeting of all NFDI4BioDiversity co-applicants and participants to provide room for peer to peer, task areas, group and team communication to discuss and exchange the overall progress of the project. It will be structured by topic-driven workshops as well as general sessions including the involvement of partner NFDIs and external speakers and guests.
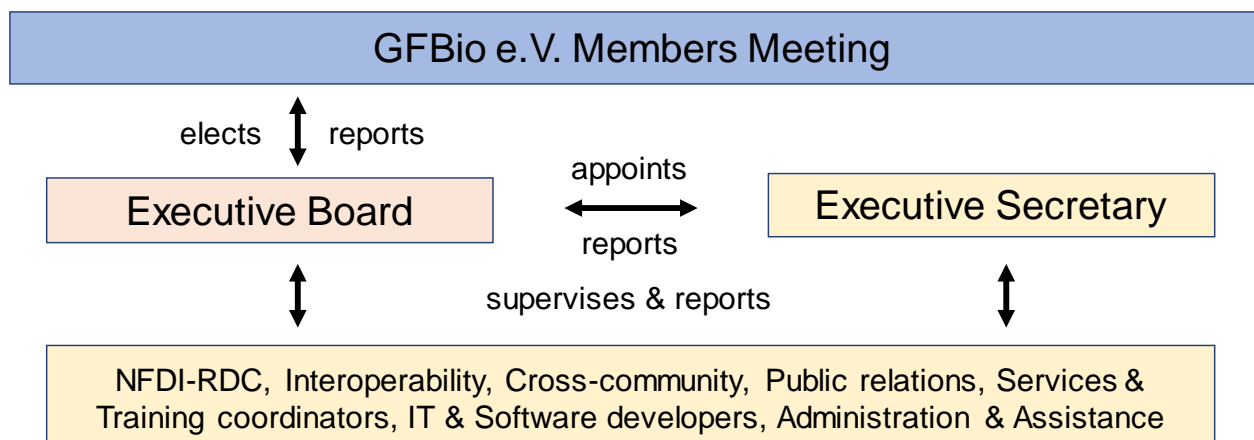
Figure 2.5.3: The structure of GFBio e.V., including the personnel directly employed by GFBio e.V. to coordinate NFDI4BioDiversity

To professionally fulfil its duties all critical core services provided by the consortium will be operated by GFBio e.V.

This includes the webpage with data submission and search functions, data management planning, the helpdesk's first level support, legal advice, quality management and certification of archives. Consequently, besides the personnel for assistance, financial and personnel administration, a set of dedicated key persons will be directly employed by GFBio e.V. (see Task Area 5 for details).

IT and Software developers: The pool of IT and Software developers will act as a task force for the flexible support of co-applicants and participants in case they encounter technical or software problems in providing or receiving data. To access this pool, any of the party have to hand in a written application describing the purpose and duration of the task that needs to be supported. The SC will review the applications within four weeks and together with the speaker it will decide if and how much person months (PMs) will be granted to the applicants. A lean application procedure is envisaged to quickly respond to any requests and get the problem solved soon. For the participants named in the application task-force, PMs are already allocated based on their requests.

Overview of NFDI4BioDiversity's governance: The Speaker, GFBio e.V., SC, SAB and BoG will work closely together to govern NFDI4BioDiversity (Figure 2.5.4).
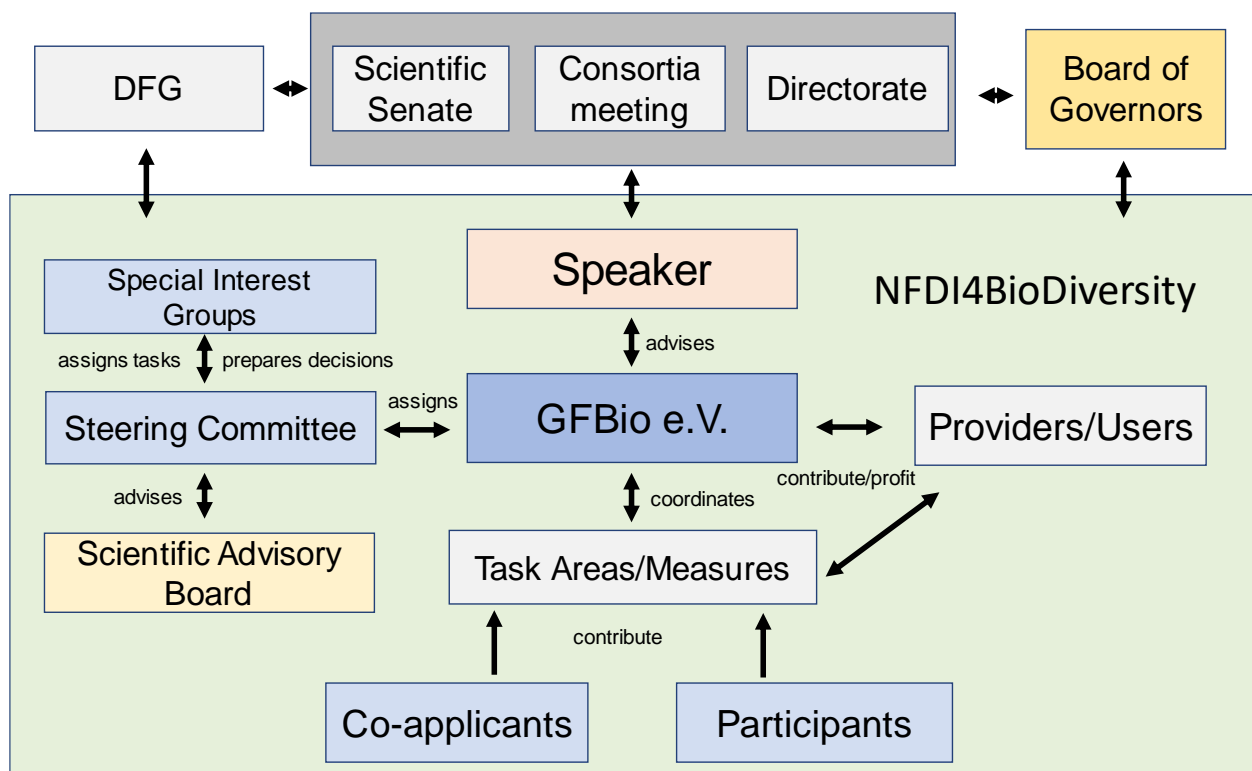
Figure 2.5.4: Overview of the different bodies in NFDI4BioDiversity and NFDI their interactions

Viability: Besides the direct involvement of anyone interested in NFDI4BioDiversity by becoming a member of GFBio e.V., instant user satisfaction surveys will be run to get the pulse of the users. They will be complemented by feedback sessions within PR/workshops and training/education activities.

Besides the above described short-term "project"-like management and coordination, tracking and controlling NFDI4BioDiversity will implement a long-term strategic management to be able to respond to substantial changes in the political, scientific and partners level (SIG5). With NFDI being developed as an infrastructure with a potential timeline of 15+ years it can be expected that PIs and partners change over time due to retirement and/or change of scope. This can be accommodated by adjusting the co-applicants in the consortium in each phase of the funding periods. In case archives or IT infrastructure partners are dropping out a careful evaluation will be undertaken if alternatives exist or if other partners can take over. In case of software, it will be evaluated if it can be maintained by GFBio e.V. Changes on the political or scientific level will be accommodated by expanding or reducing the consortium depending on the needs of the research community. In case the consortium will be decommissioned an EXIT-strategy is in place (see Task Area 5).

General means of communication: The primary method for information exchange between the NFDI4BioDiversity members is by organizing regular technical meetings, video conferences and workshops at various management and working levels (Steering Committee, Task Areas, Measures, Special Interest Groups, Cross-Cutting Topics). Reporting has been proven to be an effective means to inform and communicate the overall project progress. Therefore, TALs and MLs have to regularly provide milestone reports, all partners have to provide input to the periodic and evaluation reports. GFBio e.V. will compile reports and circulate them among the consortium members. All meetings are announced and scheduled well in advance; virtual meetings can be scheduled at shorter notice. Agendas, proposed resolutions, and working documents are made available in advance to all attendees. Responsibility for initiating the members meeting, SC, SAB, BoG meetings and interactions with the Directorate and Senate as well as the agenda, rests with the Speaker together with GFBio e.V. For task area, measure meetings and technical workshops, the TALs and MLs leader who calls the meeting is responsible for preparation.

Means for resolving conflicts and risk management

The TALs and MLs must immediately inform the Speaker of any problems in achieving their objectives, including the reporting of any potential conflict situations. The first opportunity to resolve conflicts is the SC. In case the conflict remains, the Speaker will collect the differing positions which then have to be voted on in the members meeting by a simple majority.

In case of urgency, electronic voting will be established. In case of further escalation SAB, BoG and the NFDI Directorate will be consulted.

A large-scale project such as NFDI4BioDiversity faces many risks, for example, due to its size and the yet unclear long-term sustainability of personnel and expertise. This may have considerable effects on the project progress. The following table provides the NFDI4BioDiversity risk assessment:

Table 2.5.1: Risks and Measures

Risk	Effect	Scale	Measures
Processing and/or archiving components do not scale with acquired amount of data	Processing and/or archiving tasks are delayed or cancelled	Medium	Increasing storage and processing capacities by additional commitments from NFDI4BioDiversity partners, new participants and fund raising
Curatorial capacity does not scale with submitted amount of data	Delays in archiving and publication of data; reduces confidence in the service; decrease user acceptance	Low	Increasing curatorial capacities, by additional commitments from NFDI4BioDiversity data centers
Lack of user acceptance	Project failure	Medium	Increase priorities related to PR/outreach (SIG1), training/education (SIG3) and cooperation activities with other NFDIs, users and providers
Shift of research and infrastructure requirements due to global developments	NFDI4BioDiversity developments do not match the overall objectives	Low	Monitoring requirements and adaption of the work programme providing more generic services (SIG5)
Partners do not have the same speed of progress	Delayed actions on one side could slow down speed of entire project	Medium	Monitoring progress (KPIs, reports, statistics) and shifting tasks to other partner by changing priorities (SIG2)
Diverging views on project priorities	Affects the entire project planning and progress	Medium	Get SAB, BoG and Directorate involved to discuss and possibly reset priorities (SIG5)
Partners are pursuing their individual objectives disregarding common goals	The integration of different NFDI4BioDiversity components will be affected	Medium	SC asserts reporting obligations and intervenes in time. GFBio e.V. with TALs and MLs should advise and keep track on progress within affected TAs and measures. Escalate to BoG and Directorate. As a last resort, reduce the next share of funding
Single partners drop out	Development and maintenance of components might suffer	Medium to High	In case key infrastructure components are affected GFBio e.V. will try to take over (funds will be reallocated). Ask for other partners to step in or involve additional participants or consortia to fill the gap
Key project staff leaves	Delay in progress of TA work	High	Recruit immediately alternative capacity to cover personnel loss, improvement of internal work documentation; knowledge base
Project positions cannot be filled	Work cannot be initiated by partner, delay in TA work	Medium	Establish task groups to fill gaps with partners, in-kind staff. Recruit NFDI wide in a concerted action

2.6   Operating model

Operating model

As discussed in the DFG Governance workshop in Bonn on 30.08.2019, NFDI4BioDiversity is in favour of the establishment of an overarching NFDI e.V. We are not planning to transform NFDI4BioDiversity into an independent legal entity within the next years, until this might facilitate the creation of NFDI e.V. With GFBio e.V. in place as a co-applicant, we are open to share our experiences in setting up a not-for-profit association. We recommend that the statutes of the future NFDI e.V. should be open for memberships of associations. Another option would be merging GFBio e.V. with NFDI e.V. In this case also the coordination duties of GFBio e.V. in NFDI4BioDiversity must be transferred to NFDI e.V. As indicated - we are open to any of these models and would appreciate a single legal entity in place since it would facilitate the whole legal framework as well as operating, governing and sustaining NFDI. We understand that many open questions must be clarified until a NFDI with a legal model can be established. Consequently, we will run NFDI4BioDiversity using the standard operating model for a consortium. This includes a CA between the applicant institution and all co-applicants to allow the transfer of money on purpose. Special attention will be given to the financial compensation model for participants, where no exchange of goods and services in a commercial sense is currently foreseen. Since all NFDI consortia must solve these issues, we already agreed to work together to design a CA and contracts that are compliant with German law.

Table 2.6.1: Expertise, resources and infrastructures of the co-applicants and participants available for the NFDI

Co-applicants	Expertise and Experiences	Infrastructures, Repositories, Software	References
UniBremen-MARUM	Operates the PANGAEA information system. Data management, curation and publication. Web services, indexing and search technologies, APIs, helpdesk. Expertise in certification, standards development and ontologies. ICSU World Data System and CoreTrustSeal certified, accredited by the WMO as Data Collection and Processing Center.	PANGAEA	[35, 78, 79]
BGBM	Operates the EDIT and Euro+Med Plantbase. EU BON taxonomic backbone, JACQ botanical collection management and terminology services. Established scientific collections of herbarium specimens. Data management, archiving and publication. Taxonomic computing, primary data processing, data modelling and standardisation.	Euro+Med Plantbase, EU BON taxonomic backbone, BioCASe, EDIT, AnnoSys, GGBN	[80–86]
BIBI	Coordinating and administrating the de.NBI network and the German ELIXIR node. Organizing the governance of the federated de.NBI cloud and running the de.NBI cloud. Developing microbial bioinformatics for microbial genomics and metagenomics. Educating and training bioinformatics and research data management.	de.NBI ELIXIR-Germany, EOSC-Life	[34, 87–90]
DSMZ	Operates BacDive the largest repository for bacterial phenotypic data and the Taxonomy service Prokaryotic Nomenclature Up-to-date (PNU). Expertise in identification, characterization, and description of microorganisms. Compliance with the Nagoya protocol and documentation, first registered collection according to EU regulation 511/2014. ISO 9001 certified.	BacDive, Prokaryotic Nomenclature Up-to-date	[91–96]
FBN	Operates the FarmPheno database for the genotype-phenotype-distinction in the context of farm animal performance, health, and welfare. Established research data management processes to make the data usable and reusable according to the FAIR principles and established metadata standards. Pipeline to link research data with sample repository.	FarmPheno	[97, 98]
GFBio e.V.	Expertise in research data management and community engagement. Unified submission and brokerage system for complex biodiversity data sets, feeding into workflows for data archival and publication of nine data centers (incl. APIs). Data management planning. Helpdesk.	GFBio.org	[43, 99, 100]
GWDG	Design and operation of large, distributed service infrastructures. Design, implementation and operation of data repositories, archives and data management pipelines. Consulting and training for data management planning. Virtual research infrastructures. Development and customizing of a metadata quality assessment framework. ISO 9001:2015 and ITIL certified.	IDM and SSO, Cloud server, storage, data repositories, PIDs SaaS (BEXIS2, DWB)	[101, 102]
HITS	Operates the FAIRDOMHub data management system and data commons for systems biology and the RightField tool for data annotation. Data management, curation and publication. Ontologies and spreadsheet/CSV related tooling.	FAIRDOM FAIRDOMHub RightField SABIO-RK	[103–108]

DFG

Co-applicants	Expertise and Experiences	Infrastructures, Repositories, Software	References
IPK	Operates the Plant Genomics and Phenomics Research Data Repository e!DAL-PGP and Genebank Information System GBIS. Data management, curation and publication. RDMO tool for creation of DMPs. Compliance with Nagoya protocol. Actively contributes to the development of standards for plant phenotyping like MIAPPE and programmatic interfaces to access plant diversity resources like BrAPI. ISO 9001 certified.	GBIS/I, e!DAL-PGP, EURISCO	[109–114]
SNSB	Established infrastructure as GBIF data publisher and GFBio data center (incl. archiving services, DOI assignment). Meta´omics schema development MOD-CO. Software tools to manage schemas, ontologies and taxonomies. Support user with Diversity Workbench (DWB) data management tools and workflows (DWB User Helpdesk), data repositories and portals of LIAS, DEEMY and Bayernflora, DiversityMobile App. Involved in the consolidation of the technical infrastructure of the biowikifarm.	Diversity Workbench, MOD-CO, Biowikifarm, LIAS, DEEMY, Bayernflora, Taxonomien, DiversityMobile App	[48, 49, 115–126]
UFZ	Consulting for data management plans and use of metadata standards. Use-case driven workflows and pipelines for management of heterogeneous data streams (e.g. time-series, high-throughput data, samples, field management, structured file-based data). Software and data pipelines for automated quality assurance, metadata enrichment, machine learning and other data analytics. Standards and technologies for automated metadata generation. RDM training, controlled vocabularies, ontologies.	ModWebBioDiv, eLTER RI ESFRI, HI-CAM, MOSES, TERENO, ESM, ICOS, LTER-D	[50–55, 127, 128]
UniGießen	High performance computing. Standardized software containers. Kubernetes infrastructure. Cloud computing. User support and training. Data management and automated reproducible analysis workflow development for high-throughput omics data. Software platform for microbial bioinformatics. Structured acquisition and storage of experimental data and the automated processing of that data. Governance and monitoring of the European Open Science Cloud (EOSC).	ASA3P, Conveyor, GenDB, EDGAR, MGX, ReadXplorer, EOSC	[89, 129–131]
UniJena/iDiv	Developer and provider of the BEXIS2 data management platform. Operation of BEXIS2 instances for several large-scale projects. Consultation and training on research data management leveraging semantic techniques for RDM. Provenance for scientific data and workflows. Operates the sMon project 'Trends in German species data'. Strong science-society-policy interface, incl. citizen science.	sMon, BEXIS2	[59, 132–136]
UniMarburg	Development and maintenance of the VAT system that serves for the visualization, analysis and transformation of biodiversity and environmental data. Jupyter notebook integration. Visualisation of essential biodiversity variables. Data streams (e.g. time-series) and scalable monitoring systems. Major competencies in promoting data and code literacy amongst young scientists and students.	VAT system, FOKUS, TRUST	[137–139]

DFG

Participants	Expertise and Experiences	Infrastructures, Repositories, Software	References
AraGes-Spinnen	Operates the ATLAS (ARAMOB) portal. Data ingest, quality management by experts. Minimum data/metadata standards. Voucher deposition. Education and training of volunteers and scientists.	ATLAS ARAMOB	[140–142]
AWI	Professional data management and data publication environments. Observation to Archive (O2A) software covers the full life-cycle of data from source to analysis and re-use. International networking.	O2A	[143, 144]
BGBM-DI	Operation of repositories (PhycoBank, AlgaTerra, SpecimenTool, GGBN), reusable software (BioCASe, EDIT) and pipeline (MetBan). Long-term data preservation, publication and quality management. Nomenclature and taxonomy of algae specifically diatoms. Data curation.	AlgaTerra, SpecimenTool, GGBN, BioCASE, EDIT, MetBan	[82–84, 86, 145–152]
BIOfid	Access to current and historical biodiversity literature, reusable tools for text-mining, state-of-the-art semantic search technology, ontologies and terminologies.	BIOfid's web portal	[153–155]
BSH	Web-based information network for marine biological data. Environmental impact assessment and monitoring data from offshore projects.	MARLIN-BSH	[156, 157]
BUND	Managing data from citizen sciences projects e.g. Felis silvestris silvestris or Eliomys quercinus. Education and training.		[158]
DDA-Vögel	Operates the Ortnitho web portal for bird monitoring with 41.6 million records. Coordination of four national, structured monitoring programmes with 11,000 active birdwatchers. Training and outreach, quality management, data integration and harmonization.	Ornitho	[159]
DNAquaNet	Technical criteria for DNA and eDNA metabarcoding pipelines to assess aquatic biodiversity changes. Development of relevant metadata standards for DNA/eDNA data sets together with GGBN. Training scientists and industrial stakeholders.	DNAquaNet	[160, 161]
GdO-Libellen	Community engagement and education/training. Search services and data.		[162]
Gfi-Fische	Information portal on fishes in Germany.		[163]
GfÖ	Access to the ecological community. Protection of species and genetic diversity as well as the preservation of intact ecosystems, landscapes and habitats. Outreach and training.		[164]
HIFMB	HIFMB holds expertise for data analysis for marine biodiversity data worldwide, especially with respect to statistical ecology, time series analysis and spatial analyses. Data entry into MARLIN-S biodiversity information system, including quality-control feedback loop. Continuous update of (taxonomic) information, including tracking of data history.	MARLIN-S	[157]
HLNUG	Operation of the database system MultiBase CS with 2.5 million species specific data records. Terminologies and Ontologies.	MultiBase CS	[165]

DFG

Participants	Expertise and Experiences	Infrastructures, Repositories, Software	References
IfL	Conception, editing, cartographic design and production of various thematic maps and visualizations for print and cross-media atlases and other products. Usability testing.	Nationalatlas	[166, 167]
IGB	Operates a database for freshwater biodiversity including six lakes and two river sites. Long-term monitoring (for 60 years) and analysis of biodiversity data for phyto-, zooplankton and microorganisms since 20 years.	BenGenDiv	[168–170]
InfAI	Big data and linked data management, terminology management. Education/training in semantic modelling of terminologies/ontologies and linked data knowledge graphs. Extract structured content from Wikimedia projects.	DBpedia	[171, 172]
IÖR	Operates IÖR-Monitor for the description and analysis of land use and land cover. Research data management and workflows for land use data.	IÖR-Monitor	[173, 174]
JKI	Searching repositories (e.g. OpenAgrar) and databases (e.g. gene banks, pests, antagonists, pesticide use, crop management, long term field trails). OGC services within the GDI-DE. RDMO tool for creation of DMPs.	OpenAgrar OGC services RDMO	[175]
LAU	Operation of the database system MultiBase CS. Biodiversity data of protected areas.	MultiBase CS	[165]
LfULG	Operation of the database system MultiBase CS. Biodiversity data of protected areas.	MultiBase CS	[165]
MFN	Established infrastructure for hosting and providing research data and services. Optimization of data management in research museums. Data standards and harmonized data flows facilitating documentation, exchange and archiving. Analysis, development and establishment of methods for efficient data linkage, (automated) data enrichment and annotation. Training on good practice of research data management.	Animal Sound Archive digital library Europeana, Forschungsfall Nachtigall, Naturblick App	[176–180]
MPI-Biogeochem	Technologies for exploring big data in the Earth system sciences. Interpretation of the biodiversity data in a spatiotemporal context. Data cube interfaces. Dimensionality reduction. Long-term data hosting and sharing.	DataCube Interfaces, Big gridded data	[181–183]
Naturgucker	Operates the naturgucker web portal. Exchange information and social information with other observers. Quality management of the data by social exchange. eMentoring of observers.	naturgucker.de	[23]
NetPhyD-Pflanzen	Operates the Deutschlandflora portal a database of 30 million datasets of floristic distribution data and Recorder-D for occurrence and distribution of species and habitats. Data quality management, based on taxonomic group experts and regional specialists.	Deutschlandflora	[184, 185]
NP Bayerischer Wald	Provisioning of data products from the national park.	Nationalpark App	[186]

DFG

Participants	Expertise and Experiences	Infrastructures, Repositories, Software	References
NP Hunsrück-Hochwald	Operates the National Park Research Server. Building and managing metadata stock. Connecting and management of distributed and cascading metadata stocks. Interoperability.	National Park Research Server INGRID	[187, 188]
SAB	Provisioning of research data from Bavarian agricultural, environmental and other agencies. OAIS-conform digital long-term archiving and access provision for complex data ensuring their integrity and legal validity.	Findmitteldatenbank	[189]
SGN	Operates biological collection databases, vegetation and species occurrence databases as well ecological and citizen science databases. Submission/search services, education/training, terminologies/ontologies, data products.	IPBES DiSSCo prepare Insekten Sachsen	[58, 190–192]
SMNS	Established infrastructure as GBIF and GFBio data publisher. Operates the Naturportal-suedwest.de, a portal for species identification (gramineous plants, urban insect fauna, terrestrial and limnic molluscs, and fossils) with tools for the broad public to report species observations. Data management, curation, archiving and publication. Mass digitization. Data quality control. Terminologies, ontologies.	Naturportal-suedwest.de	[193]
SUB	Expertise with numerous strategic projects and standardization activities, as a leading partner (KIM, LIDO, TEI, Dublin Core, RDF Application Profiles, MODS, COAR, OpenAIRE, RDA, DINI). Training and education (FOSTER, FIT4RRI, FAIRsFAIR). Open access information platform open-access.net. Intellectual preservation of information, long-term storage, digital editions and indexing of data.	open-access.net	[194, 195]
Thünen-GDI	Operates the TISDAR database for the provisioning of datasets from the federal monitoring like 'Bodenzustandserhebung' and the national biodiversity monitoring. OGC conform services like WMS, WFS, and CSW.	TISDAR	
Uni Göttingen	Establishing case studies and corresponding educational resources based on research data. Facilitating community engagement.		[196]
UniLeipzig	Access to existing data information infrastructures in the limits of existing data sharing agreements and HPC cluster Coordination of the planned PlantHub at iDiv, providing access to a wealth of plant related data		[197]
ZFMK	Established infrastructure as GBIF and GFBio data publisher. Operates the MorphoDBase database for the anatomy of organisms and the German Barcode of Life Project Webportal. Data workflows for genetic barcoding and transcriptome data management. Long term data preservation, publication and quality management for anatomical and occurrence-based collection- and research data.	MorphDBase, German Barcode of Life Project Webportal	[198–200]

DFG

Participants	Expertise and Experiences	Infrastructures, Repositories, Software	References
ZMT	Access to tropical marine biodiversity data. Evaluation of heterogeneous, discontinuous, and large data sets. Implementation of comprehensive institutional data management practices.		[201]

DFG

3   Research Data Management Strategy

Current state of research data management in biodiversity

In Germany, mostly natural history collections and museums have the expertise and infrastructure to manage research data in a sustainable and organized way. As digitalisation on the provider and user side proceeds, they lack capacities to cope with the flood of heterogeneous data from multidisciplinary projects. This is essentially similar to the situation on the global level, leading to the existence of numerous isolated, and often technically proprietary and unsustainable databases and "repositories". A coordinated approach, similar to the Australian National Data Service (ANDS) [202], is needed to improve the current situation. To specify stakeholder requirements, the DFG established a working group (AG Biodiversitätsdaten) in 2011 focusing on the handling of biodiversity data and the investigation of user requirements for the sustainable use of research data. As a result the AG published a strategy paper on the treatment of environmentally related biological data [203], recognizing four major challenges within the current national e-infrastructure landscape: 1) Deficiencies in the organization and sustainable funding of existing archive infrastructures, 2) Lack of interoperability due to insufficient and/or incompatible standards, 3) Fragmentation of existing service infrastructures comprising data acquisition, publication, retrieval, dissemination, and integration 4) Lack of awareness and skills about data acquisition, management, and submission in the research community.

In summary, the results stipulated a coordinated strategic approach for RDM in biodiversity research, which resulted in the GFBio project. During the last seven years, GFBio designed and implemented a coherent infrastructure for data acquisition, curation, archiving, publication, and integration of FAIR high-quality data products for biological and environmental sciences. Already before the idea of the NFDI arose, it was clear that not only the expertise, innovations, and resources of existing biodiversity and IT infrastructures are needed, but stronger participation of the biodiversity communities in the overall development process is essential. This has been realised by assembling the NFDI4BioDiversity consortium.

Consortium specific expertise and experiences, existing information infrastructures, data repositories or reusable software

The consortium provides a wide range of expertise, experiences in RDM and user support interactions, and a wealth of infrastructures ranging from data repositories to computing and software services. Detailed information about the contributions and roles of co-applicants and participants are given in section 2.2, Table 2.2.1: Summary of NFDI4BioDiversity's participants, their roles and contributions, section 2.6, Table 2.6.1: Expertise, resources and infrastructures of the co-applicants and participants available for the NFDI, and in section 6 (Co-) Applicant Contributions.

Envisioned state of research data management

NFDI4BioDiversity builds on the resources, services, and expertise of members of GFBio, de.NBI pus additional co-applicants and 35 participants with their specific expertise and services as summarized in Table 2.6.1. Together we will substantially improve biodiversity related data management by the following measures:

1. The significant engagement of researchers, authorities, societies, associations and citizen scientists involved in prominent biodiversity related projects and initiatives. For this purpose:

- we selected a number of use cases representing the spectrum of biodiversity research, monitoring, authorities, associations, agencies and citizen science. They signalled a strong need to improve their current data management approach, adapt their workflows to the proposed NFDI-RDC and include data science (TA1 M1). Some of the use cases are cross-domain and embedded in several NFDI consortia. SIG1 (public relations) will take care that seamless communication between the consortium and the community is reached.

- we will provide support with usage, adaptation and deployment of tools and platforms for gathering, structuring and (early) mobilization of data, as well as for project data management, analyses and education. A focus will be on tools and platforms supporting collaborative work and data science (TA1 M2). Further on, we emphasize that TA1 M2 will concentrate on direct technical advice and community support with generic (if possible) and widely used tools to process and structure data to improve FAIRness (TA4 M3).

- we will provide data management and data science support for researchers, research groups, science projects and our community by co-design on an "as needed" basis (TA1 M3). A helpdesk will support users in all stages of the data life-cycle. Direct contact with our community is essential and requires skilled domain experts. Particular attention will be given to the data management support for our use cases (TA1 M1).

- we will organize an education and training platform to promote RDM and data literacy (TA1 M4). The research background of our use cases (TA1 M1) will thereby deliver the scenario for a number of case studies. Resulting materials will be gathered, evaluated, and prepared for further usage in associated universities and NFDIs. The education and training platform will help embedding RDM and data literacy in universities' curricula and will contribute to the needed cultural change. SIG3 (training) will coordinate the consortium's activities.

- we will monitor the effectiveness of our services and community engagement by a number of activities (TA1 M5): 1. regular surveys and addressing researchers in our use cases and the biodiversity communities in general, 2. user workshops, 3. collection of user feedback, and 4. monitoring the quality of supplied services (TA4 M5). Results will be used to iteratively improve the quality and effectiveness of our measures.

- User satisfaction will be the guiding principle of NFDI4BioDiversity. SIG2 (service monitoring) will coordinate the activities across the consortium.

- we will interact with our community ranging from scientists to volunteers to first gather a common understanding of "data quality" as a deduction of application. This will guide our data selection and curation process with respect to FAIRness, "fitness for use" and "fitness for purpose" (TA1 M5). See also section 3.2.

Community engagement (Task Area 1) has a significant share of the overall budget (~36%). The investment is intended to leverage a process towards more awareness, data science and collaborative work.

2. The NFDI Research Data Commons (NFDI-RDC) (Task Area 4), as an overarching virtual and expandable infrastructure to leverage user involvement and collaborative data-driven research (see Figure 4.5.1, TA4 M4). The NFDI-RDC concept includes joint cloud services, an application framework with access to computing power and collaborative workspaces, and a common authentication and authorisation infrastructure (AAI) (TA4 M5). **The NFDI-RDC creates a bridge between data providers and users.** It is expected to have a significant effect on the data quality and openness, as well as the efficiency of using existing data and thus supplies a strong incentive for users. For the implementation of the NFDI-RDC we will:

- create interoperability and workflows with national and international data repositories & services (TA3 M3, TA4 M1) as well as with further NFDI consortia (TA2 M1). To reduce complexity and efforts needed for development and maintenance of the resulting broker framework we will push for a limited number of protocols and content standards (cross-cutting topic). One of the favoured champions on the metadata level is schema.org which comes with a community extension for biodiversity. More details are given in section 3.1.

- integrate and harmonize data from repositories and data services (TA4 M2) with a focus on the development of data products (TA3 M3). The measure will address structural, semantic, and conceptual problems by 1) developing and adopting common (meta)data models, 2) extension and application of the GFBio terminology service, and 3) adaptation of tools for data transformation and harmonization. The latter will leverage work on data integration as well as schema and ontology matching and include machine learning and AI approaches and knowledge graphs. Further on, to enable cross-domain data exchange, we will communicate and negotiate our (meta)data models with all NFDI consortia (TA2 M1). This organised integration and harmonization work provides the most significant gain in efficiency for the preparation and compilation of data for research applications.

- develop approaches to integrate streaming data, a data type that is rapidly gaining importance in our domain. In the NFDI-RDC this will be built on approaches from the data streams community and solutions like the Earth Data Cube (TA3 M3).

DFG

- consolidate harmonized and integrated data into cloud-based semantic storages as backends furnished with APIs which allow general access and efficient usage of data and associated services (TA4 M3). Integrate new types of standard data products structured according to existing and evolving community accepted data and metadata standards (TA3 M3).

- supply the building blocks for an application framework (TA4 M4) including elastic compute resources, collaborative workspaces, components to deploy and share tools and code, and an education and training platform (TA1 M4). Besides supporting specialized community portals, we will extend the current GFBio data portal comprising metadata from related NFDI consortia.

- Internationally, the development of Research Data Commons is still at an early stage. Examples are the RDC of the US National Cancer Institute [204], and the Australian Research Data Commons [205] which we will collaborate with. Nationally, we will collaborate with all NFDI consortia, in particular NFDI4Earth and NFDI4Agri, to shape and create the NFDI-RDC (Berlin Declaration). SIG5 (strategic planning) will be in charge to coordinate the strategic activities within the consortium and across-NFDIs. We regard the NFDI-RDC as an important building block for the development of ONE NFDI.

In summary, NDFI4BioDiversity envisions improving and innovating RDM in a two-phase approach. In the establishment phase (2020-25) we will reach consensus and consolidate the services and demands of our co-applicants and participants as part of the biodiversity community. All strategic decisions and developments will be designed for embedment in NFDI. In the second phase (2025-30) stable and operational support of RDM from citizens to researchers is envisioned. We expect NFDI4BioDiversity services to support the full data life-cycle and to facilitate data science and data analytics for multidisciplinary research, innovation and decision making. To fully unfold, this requires clear and stable overall governance structures as well as a clear legal framework for NFDI as a whole. The applicant and all NFDI4BioDiversity co-applicants have agreed to foster open science and open data.

3.1 Metadata standards

Biodiversity research has become truly multi- and interdisciplinary: progress in the field now depends on collaboration across a wide range of scientific disciplines to understand and monitor living systems from molecular to planetary scales. However, most of these disciplines have evolved in relative isolation from one another for decades, and still use remarkably localised standards and terminologies to describe their data, information, and knowledge.

In the field of biological primary data, the TDWG Biodiversity Information [7], DwC [206] and ABCD [207] standards are widely used. ABCD, in particular, is used by the natural history data archives involved in GFBio for the storage of collection and observational data [208].

Other relevant community data standards include EML (Ecological Metadata Language) [209], MIxS (Minimum Information about any (x) Sequence) [210], SDD (Structured Descriptive Data) [211] and the GGBN Data Standard [151, 212]. In addition, various standards have been established or adopted to ensure uniform access interfaces for biodiversity data (e.g. OAI-PMH [213], BioCASe-Protocol [214]) or to deal with data annotation (e.g. W3C Open Annotation Specification [215] and CETAF Identifier [216]). Evolving generic conceptual schemas, like the MOD-CO schema [50], guide scientists in new research fields for processing sample data in meta-omics research and facilitating data transformation.

Talking a common metadata language

Several approaches have been initiated by TDWG to integrate geospatial and biodiversity data and by GEO BON to integrate data from different domains. Domain specific approaches include the definition of complex namespaces and schemas such as those used by the DiversityWorkbench [217] or the EDIT Common Data Model [84] for mapping biodiversity and ecological data. Efforts have been made to map relevant data exchange standards such as ABCD [218], DwC [219], ISO19136 [220] or the GBIF metadata profile [221]. On the European level, the INSPIRE initiative [222] supports the ISO19139 schema [223] implementation. Emerging standards are now based on structured data such as BioSchemas [224].

Despite the progress made by these groups, there is still a considerable mismatch between the different metadata standards. This hinders interoperability and a unified way to search biodiversity resources (occurrences, species, datasets). Through the consultation with users in Task Area 1 we have and will continue to identify the minimum metadata elements that are required, e.g., to support the usage and registration of biodiversity resources and determine relevant content metadata standards. Design considerations such as clarity, flexibility, and extensibility will be considered when choosing the relevant standards. schema.org is one of the champions having the potential to integrate metadata cross-domain. Schema crosswalk between minimum metadata across domains is vital to facilitate the development of NFDI-wide metadata standards. Activities such as standardization and schema mapping will be fostered through Task Area 2. The rich participation of our partners in the relevant standardisation organisations ensures that international developments are incorporated into the NFDI process and, conversely, that specific requirements are incorporated into the international processes (TA2 M3).

DFG

Semantic integration: Ontologies

Technologies based on knowledge representation (KR) offer one solution to issues stemming from heterogeneous terminology across biodiversity research communities. KR approaches encoding varied knowledge in logically consistent and machine-actionable semantic ways is known as "ontologies". Fortunately, the biological and biomedical domain is home of the most stable and successful KR federations: the Open Biological and Biomedical Ontology (OBO) Foundry and Library [225]. OBO resources and know-how - alongside its inclusive editorial policies - are well-placed to support the NFDI4BioDiversity mission, particularly as the federation has recently expanded to include ontologies handling ecological and environmental entities. In addition, other notable groups are creating shared vocabulary and thesauri services relevant to biodiversity research. Examples include ESIP [74], GBIF [3] and LTER [8].

This ontological landscape - while already usable - requires user communities to shape its content and structure such that it can meet "on-the-ground" needs. Through the NFDI4BioDiversity consortium, we will engage regional biodiversity researchers and service providers to champion their needs among the developers and federations of semantic interoperability systems. With several "ontology" partners on board, we will leverage existing collaborations with key developers and representatives of global reference ontologies to achieve rapid and practical advancements to prevent new silos from forming. The impacts here will promote the use of NFDI4BioDiversity services in both a national and international context and promote interoperability with other systems through coordinated use and co-development of global semantic infrastructures propagated by Task Area 2.

The semantic data integration within the NFDI4BioDiversity infrastructure and with external partners requires a reliable service that provides the NFDI relevant terminologies and ontologies in a stable and performant way. The Terminology Service developed as part of the GFBio initiative [62], provides a unified API for cached terminologies required in processes such as data integration, annotation and query expansion. TA4 M2 will expand the service with a focus on versioning and synchronization of terminologies, performance of decentralized semantic resources and effective and transparent workflows for coordinating terminologies required in research processes. The terminology service will thus be expanded into a broker system for semantic applications in NFDI4BioDiversity and beyond. An important application will be the services provided by Task Area 4 for the transformation of data schemas, whose semantic basis (terminologies, ontologies, relationships between concepts) can be obtained from the terminology services.

DFG

3.2   Implementation of the FAIR principles and data quality assurance

The FAIR principles are high level concepts; therefore, their concrete implementations are left to the users. The principles do not explicitly define the quality aspect of meta(data). NFDI4BioDiversity will translate the FAIR principles into practice while ensuring that minimum requirements for data quality (e.g. completeness, consistency and availability) and curation (e.g. documentation of system and services) are achieved. In order to meet the community requirements, the technical solutions enabling data FAIRness and quality will be implemented and tested incrementally based on requirements and feedback from Task Areas 1 and 2. SIG2 (service monitoring and FAIRness) will coordinate the activities. Sustainability of the solutions will be ensured by designing the solutions following the governance and sustainability guidance outlined in Task Area 5. The following technical solutions and activities will be implemented:

- (Early) data mobilization (long tail, citizen science data, data from authorities and societies and hidden data from collections), supported through TA1 M1, M2 and TA3 M2 which produce guidelines, uniform workflows, and offer technical and curatorial support based on open source tools, standards and communication protocols.

- Tools supporting consistent transformation of meta(data) into widely recognized standards in biodiversity research. Appropriately transformed meta(data) will be enriched with terminological resources (TA4 M2) to improve their discovery (TA4 M4). We will gather qualified references to related datasets from data providers (through the broker framework, see TA4 M1) and infer additional data linkages through innovative approaches (e.g. text analytics).

- Data interoperability, supported by structural and semantic harmonization of data holdings in NFDI4BioDiversity affiliated repositories (TA3 M3, TA4 M2). This includes work to address the integration of research data products arising from the management of biodiversity-specific new data types.

- Participation in the GoFAIR Implementation Network BiodiFAIRse [226], or operating an own Network (TA3 M2). A further activity is the new "RDA WG InteroperAble Descriptions of Observable Property Terminology (I-ADOPT)" [227], addressing harmonization of measurement and observation types.

- Educational activities, such as summer schools & training, and fosters practices of FAIR data and research data management as part of current curricula (TA1 M4 and cross-cutting topic).

- Incentives and reputation for data producers, to support the cultural change (mostly bottom-up), in particular pushing the development of data publications as acknowledged scholarly work, e.g. via the RDA Interest Group on Data Publishing [228] and various

related working groups, latest: RDA WG data fitness for use [229] and RDA FAIR data maturity model WG [230] and AGU initiative Enabling FAIR Data [231].

- Certification of the FAIRness of repository services and data holdings, preferably by CoreTrustSeal [66]. This is already part of the GFBio work program and the major objective of the H2020 project FAIRsFAIR [67] (NFDI4BioDiversity partners participating). Activities (TA3, M2) are in line with the EOSC perspective (FAIR Data Action Plan) [232].

3.3   Services provided by the consortium

Besides the services already provided by the applicant, co-applicants and participants, NFDI4BioDiversity will supply integrative, NFDI-RDC compliant tools and services, including but not limited to:

- Front-office/back-office - helpdesk
- Education & training platform, NFDI-RDC support, user forum
- Tool evaluation and tool deployment service (technical user helpdesk)
- Data quality instruments (e.g. BioCASe harvesting feedback reports, OpenRefine extensions)
- Provision of basic & initial toolset and data transformation and validation tools (GitHub, Jupyter hub, VAT system, DWB, BEXIS2, BioCASe provider software, extended GFBio services)
- Provision of collaborative workspaces including support for scientific workflow and provenance management
- Data submission, versioning and, publication service (extended GFBio service)
- Data transformation and validation tools (DWB, BEXIS 2, BioCASe provider software)
- Tools to support integration and harmonization of data and compilation of data products
- Long term data preservation and publication including DOI registration (using GFBio compliant archives)
- Search portals and APIs for data and tools (extended GFBio service)
- Terminology service (extended GFBio service)
- NFDI-RDC governance, monitoring, & support
- Elastic compute service
- Various GFBio-supported tools and de.NBI tools, see also bio.tools [233]

4   Work Programme

4.1   Overview of task areas

Task Area	Measures	Responsible Co-Spokesperson(s)
1. Community engagement (2involve)	1. Use cases to engage biodiversity communities in mobilising and archiving existing and newly generated data 2. Support with tools/platforms for (early) mobilization of data and data analyses 3. Front office/back office: Active Research Data Management (RDM) & data science support 4. Education and training 5. Regular quality surveys, user feedback, quality management, reputation systems	Aletta Bonn (UniJena/iDiv) Birgit Gemeinholzer (UniGießen) Mark Frenzel (UFZ) Klaus Wimmers (FBN)
2. National and international networking (2connect)	1. Collaboration with related NFDI consortia 2. Context with international services & developments 3. Engagement in European and international standardisation activities	Anton Güntsch (BGBM) Ramin Yahyapour (GWDG) Jens Stoye (BIBI)
3. Long-term data, tools and service preservation, certification (2consolidate)	1. Identification and description of the current and emerging technical ecosystem in biodiversity research 2. Profiling biodiversity tools, technical services and data centers, adopting them as NFDI-RDC-recommended and guide them towards certification 3. Adaptation, Enhancement and Consolidation of NFDI4BioDiversity tool and technical service backbone	Birgitta König-Ries (UniJena/iDiv) Dagmar Triebel (SNSB)
4. Data integration, exploration, and exploitation - the NFDI-Research Data Commons (NFDI-RDC) (4all & 4future)	1. Linking data and service providers with the research data commons 2. Data integration & harmonization 3. Storage infrastructure 4. Software application frameworks 5. Governance & monitoring	Bernhard Seeger UniMarburg) Alexander Goesmann (UniGießen) Uwe Scholz (IPK) Michael Diepenbroek (UniBremen-MARUM) Wolfgang Müller (HITS)
5. Coordination, Collaborative governance & sustainability	1. Coordination and collaborative governance 2. Sustainability and business model	Frank Oliver Glöckner (UniBremen-MARUM) Ivalyo Kostadinov (GFBio e.V.) Jörg Overmann (DSMZ)

All task areas are considered relevant for other NFDIs since they are the basis for the NFDI-wide cross-cutting topics. Especially in Task Area 1 we have joint use cases with other NFDIs and the NFDI-RDC (Task Area 4) should become a cooperative action across NFDI.

DFG

Dependencies among task areas:
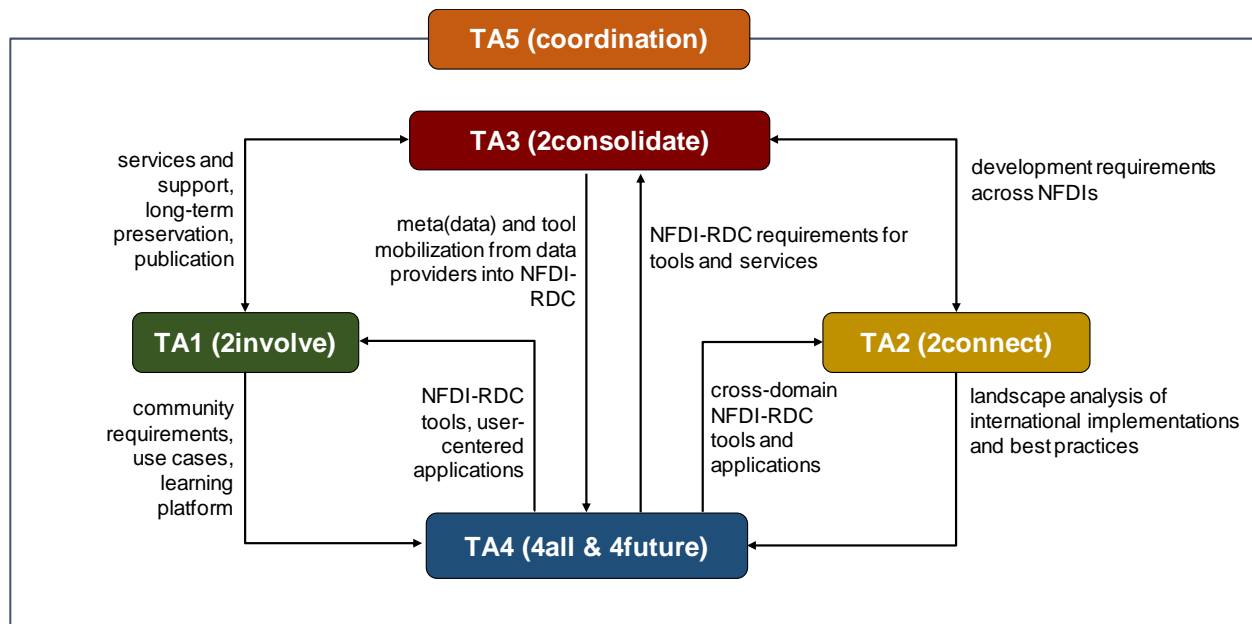
Figure 4.1.1: Graphical representation of the task areas and their dependencies and major measures.

Risks of implementation

The means for resolving conflicts and risk management has been outlined in Table 2.5.1: Risks and Measures in section 2.5. This applies for all task areas below.

4.2   Task Area 1: Community engagement (2involve)

Lead: UniJena/iDiv; **co-lead:** UniGießen, FBN, UFZ

Aims & overview: In this task area we will take care of any data management requests related to biodiversity data by building on and extending the services provided by the GFBio project. Compared to GFBio the engagement of data users and providers will be broadened by various measures and incentives like 23 use cases addressing the needs and contributions of specific communities to NFDI (TA1 M1). High-quality, innovative, user-specific tools, data products and workflows will be provided, along with web portals and services that are suitable to be widely adopted and used (TA1 M2 - in collaboration with TA4 M4). Close collaborations, awareness building and exchange across communities by training and education activities (TA1 M4) with instant quality feedback loops (TA1 M5) will actively promote the cultural change to render data management, data literacy and the FAIR data principles an integral part of any biodiversity study.

TA1 M1: Use cases to engage biodiversity communities in mobilising and archiving existing and newly generated data

Lead: UniJena/iDiv; **co-lead:** UniGießen, UFZ

Contributors: All co-applicants and participants participating in use cases

Description: Use cases are a central tool in NFDI4BioDiversity to engage with and involve representatives and major players in the biodiversity communities: 1) natural history associations, learned societies and NGO's (non-academic community, citizen science projects), 2) academic research institutions, 3) public authorities at the federal and national scale, including National Parks. Besides raising awareness for research data management in general, the main goal of this measure is to create interfaces and optimize interoperability between different data infrastructures. By close interactions with the communities we have sensed strong demands concerning data support, data mobilization and networking. Requirements differ due to different data and technical maturity and skills. As such NFDI4BioDiversity will support data standardization, implementation of ontologies, IT and software development and provision of access to NFDI4BioDiversity tools and services to use cases. The underlying strategy builds on two pillars tailored to our use cases: 1) NFDI4BioDiversity will provide meaningful support to the community to manage their data according to NFDI required standards (e.g. in terms of database structure, taxonomic backbone, metadata). 2) In return, these communities establish work- and data flows (at least metadata) with NFDI4BioDiversity considering legal restrictions and provide their input to the co-development of a national thematic network on biodiversity data for joint analysis on trends and drivers.

Our 23 use cases involving 35 participants and co-applicants, guarantee a close interaction with a broad range of communities by: 1) creating scientifically sound, high quality data sets, 2) providing examples to apply NFDI4BioDiversity tools and services (e.g. taxonomic backbones), 3) providing access to NFDI4BioDiversity analysis and visualisation tools, 4) establishment of data management workflows, 5) providing community specific web portals and services based on the NFDI4BioDiversity toolbox.

Through annual community meetings and four expert workshops per year co-organised with and designed for our community, NFDI4BioDiversity will create a thematic network on biodiversity data in Germany in close collaboration with ongoing developments in policy and practice. Furthermore, we intend to establish an instance of the ALA (Atlas of Living Australia) [234] open source software framework (see use case 23) as a powerful background infrastructure for linking existing community driven and/or citizen science data portals related to biodiversity in Germany in close collaboration with TA1 M2 and TA4 M3, M4 and SIG1.

Table 4.2.1: Overview of NFDI4BioDiversity's Use Cases

ID	Partners	Title	Keywords	PM	Year
1	UFZ, SGN	Linking biodiversity data to the European scale research infrastructure eLTER	eLTER RI (Integrated European Long-Term Ecosystem, Critical Zone & Socio-Ecological Research Infrastructure); integrate, harmonize and store biodiversity datasets; thesaurus; ontology	18	3-5
2	IPK	Diversity of plants - Molecular passport data for plant genetic resources	Gene bank access and connection; genotype-phenotype connection; plant traits	24	3-5
3	ZFMK, UniBremen-MARUM	National Biodiversity Monitoring Hub	AMMOD (Automated Multisensor Station for Monitoring of BioDiversity) station; AMMOD cloud; thesaurus; ontology	15	1-5
4	ZFMK, SNSB	German Barcode of Life (GBOL)	DNA barcoding; species habitat; species trait; taxonomic checklists	15	1-5
5	HIFMB, AWI, BSH, UniBremen-MARUM	Marine sample-based & geo-referenced organism biodiversity data	MARLIN - Marine Life Investigator; national open information system; MARLIN-S system (Marine Life Investigator - Science)	22	1-5
6	IÖR, UniBremen-MARUM	Land use/cover monitoring	OSM data; IOER-Monitor; time series; small scale; standardization of land use classifications; development of advanced indicators	11	1-5
7	DNAquaNet, BGBM-Di, UniGießen	Sample-based geo-referenced database, tools and services for metabarcoding analyses of freshwater ecosystems	Information system; stakeholder access; best practice routines; taxonomic backbone; macroinvertebrates; diatoms; fish	20	1-5
8	Naturgucker, UniGießen	Web portal "naturgucker.de"	citizen science; information portal; nature objects; data visualization; smartphone apps	14	1-5
9	SGN, SNSB	INSEKTEN SACHSEN Monitoring insects; Linking historical data with Citizen Science	Webportal Insekten Sachsen; insects; citizen science	11	1-5
10	Thünen-GDI, UniBremen-MARUM, IPK	Developing an interface between NFDI4BioDiversity and NFDI4Agri	Linking NFDIs; data harmonization; data standards; interoperability between NFDIs; ontology	11	1-5
11	LAU, UFZ, UniJena/iDiv	Linking regional and national authorities to the NFDI4BioDiversity network and *vice versa*	Information system MultiBase CS; interface; interoperability; standards; ontology	6	3-5
12	NP Bayerischer Wald, UFZ, UniJena/iDiv	Multitaxon data from a repeated sampling along a 1100 elevational gradient	Geo-referenced multitaxon sampling; protected areas	11	1-5

ID	Partners	Title	Keywords	PM	Year
13	HLNUG, UniGießen	Networking of protagonists in nature conservation	Information system MultiBase CS, common species lists, species references, field references	12	3-5
14	BGBM-DI, BGBM, UniGießen	Enabling the AlgaTerra information system for the connection to tools and services for e.g. web-/cloud-based metabarcoding analyses	Non-marine micro algae; sequence information; diatoms; biodiversity assessment; taxonomic, molecular, ecological data; automated BLAST search	6	3-5
15	GdO-Libellen, UniJena/iDiv	Connecting natural history societies to the NFDI4BioDiversity network and *vice versa*	Dragon flies; Odonata; data standardization; ontology; citizen science	16	2-4
16	AraGes-Spinnen, SMNS, SNSB, UniJena/iDiv	Connecting the data of the natural history society AraGes e.V. to the NFDI4BioDiversity network with a German monitoring portal and vice versa	Spiders; Archnidae; atlas of European spiders; Diversity Workbench	6	2
17	NetPhyD-Pflanzen, UniJena/iDiv	Network Phytodiversity Germany – connecting to NFDI	Web portal; citizen science	12	2-4
18	GfI-Fische, UniJena/iDiv	Connecting German Ichthyological Society to the NFDI4BioDiversity network and vice versa	Digital Fish Species Atlas of Germany and Austria standardization; ontology; software framework 'Biodiversity warehouse'	12	3-5
19	IfL, UniJena/iDiv	Geo-Visualization tools for biodiversity data	Spatial visualization; maps; citizen science; web map application; individual user groups	12	2-3
20	NP Hunsrück-Hochwald, UFZ	Research server for national parks	Metadata information system; documentation of research projects; geoserver and geoweb services	11	1-5
21	LfULG, UniJena/iDiv	Networking of stakeholders in nature, improving quality and readability of databases	Information system MultiBase CS; standards for collection and documentation of species observation data	12	2-4
22	MPI-Biogeochem, UniLeipzig, UniJena/iDiv	Leveraging NFDI for the iDiv community – Leveraging iDiv Data for the NFDI community	PlantHub initiative; TRY database; global vegetation plot data sPlot; Leipzig Catalogue of Plants; Earth System Data Cube; Research Data Commons	14	1-4
23	UniJena/iDiv, BGBM	Thematic network on biodiversity data to scope ALA instance	Development of thematic network, set up platform (based on ALA) for user, Research Data Commons	14	1-4

PM: Personmonth

Detailed description of the Use Cases:

1. Linking Biodiversity data to the European Scale Research Infrastructure eLTER

Contributors: UFZ, SGN, Integrated European Long-Term Ecosystem, Critical Zone & Socio-Ecological Research Infrastructure (eLTER RI) partners

Challenge: To overcome the general issue of heterogeneous biodiversity data, the EU research infrastructure eLTER RI will establish a comprehensive IT infrastructure and workflows to integrate a wide range of biodiversity and ecosystem data. By linking the NFDI to European scale RIs, the visibility and generation of integrated added value data products will increase.

Current state: eLTER RI is now starting to provide a common framework on integrating biodiversity data from a wide range of LTER sites and LTSER platforms aiming to quantify and attribute variability and long-term trends in biodiversity for different terrestrial and freshwater ecosystem types and organism groups across Europe.

Objectives & Milestones: (1) analysis and alignment of data needs for different stakeholders and data users (e.g. eLTER RI, GBIF, Natura2000, research communities); (2) identification of key biodiversity data sources (e.g. biodiversity databases, citizen science observations); (3) evaluation and implementation of procedures to align and harmonise taxonomic data from the different data sources with core taxonomic reference lists. (4) evaluation and fostering the use of common standard vocabularies; (5) evaluation and design of workflows to track provenance of data transformations; (6) establishment of workflows to integrate, harmonise and store biodiversity datasets for different taxonomic groups using trusted repositories on national (e.g. GFBio) or European (e.g. eLTER RI BioDiv component), or global scale (e.g. GBIF).

2. Diversity of plants - Molecular passport data for plant genetic resources (PGRs)

Contributors: IPK

Challenge: Create hubs for molecular passport data for PGRs to compare and analyse genotyping data for biodiversity studies or breeding schemes and link phenotype to genotype. The proposed hubs for molecular passport data are a substantial part of the long-term transformation of genebanks all over the world to bio-digital resource centers. Such a hub would focus on a single organism and would be scalable and could be expanded to include either different plant species or other data domains (-omics data). A connection between molecular passport data hubs and provider of soil and weather data (NFDI4Earth) as well as agronomic data (NFDI4Agri) would be beneficial in prediction of agronomical traits.

Current state: PGRs are preserved in gene banks like the ex-situ collection of the IPK or of WUR (see letter of support 16). These collections represent the diversity and are rich in rare alleles. Correlating genotypic to phenotypic data can give information about possible targets for crop plant improvement. Web-based systems to explore such data, like for bread wheat (Genebank 2.0) are valuable for users. One advantage but also a major drawback of these systems is the focus on

PGRs. A solution might be uploading data into such web-based systems with BrAPI (Breeding API), analysing the data and then downloading the results. Such hubs of molecular passport data should always be connected to phenotypic data to be able to make meaningful agro-economical predictions.

Objectives & Milestones: (1) Implementation of additional analyses and search functionality; (2) Development of an uploading system to link genotype and phenotype based on passport classification including error detection; (3) Programming BrAPI which integrates AAI; (4) Implementation of tools suitable to help users to make agro-economic decisions.

3. National Biodiversity Monitoring Hub

Contributors: ZFMK, UniBremen-MARUM, AMMOD (Automated Multisensor Station for Monitoring of BioDiversity Consortium)

Challenge: The use case will accompany and complement the AMMOD project (development of AMMOD stations) and ensure the interoperability of the components and software in the AMMOD project with the developments of relevant use cases in NFDI4BioDiversity. For this purpose, standards and central components will be developed that deliver the building blocks and templates for connected NFDI projects.

Current state: In various projects data are already collected over a long period of time. However, there is no large-scale, long-term and taxonomically representative automated biodiversity monitoring. This is where the AMMOD project comes in, which will use new technologies to set up automated sensor stations to monitor the diversity of species at the same locations in equal intervals. With the pilot implementation of the stations, we will obtain data on (1) time and species site occupancy rate and (2) species abundance, which will allow ecologically sound interpretations and modelling.

Objectives & Milestones: (1) Development of a catalogue of requirements for data integration of monitoring data to enable harmonization of the components; (2) Development of a thesaurus/ontology for biodiversity monitoring and further development of EnvThes ontology; (3) Integration of Citizen Science projects from NFDI4BioDiversity by expanding the AMMOD Data Cloud and vice versa.

4. German Barcode of Life

Contributors: ZFMK, SNSB, German initiative Barcode of Life (GBOL) Consortium

Challenge: Connection and expansion of the GBOL infrastructure to the NFDI-RDC so that data can be used in larger contexts.

Current state: GBOL has been successfully implemented in the last eight years. It provides a working DNA barcode reference library for animals and plants in Germany that has been established and maintained by a network of experts. In Europe, large scale initiatives have been launched in the last two years to provide reference libraries for certain faunas and floras in Austria (ABOL, initially € 500,000), Norway (NorBol,> € 10 million), Sweden (SweBol, € 750,000),

Finland (FinBOL, € 700,000), Switzerland (SWISSBOL, initially € 300,000). Together, we are building an information infrastructure that will enable a comprehensive inventory based on DNA barcode surveys, considering all types of organisms. The data is stored in DiversityWorkbench as a central backbone and becomes available via various portals and the Canadian BOLD. Taxonomic checklists for Germany (including Red List and FFH species) have been set up in the GBOL project and barcoding pipelines and IT infrastructures were established. In total, there are 24,000 types with barcodes.

Objectives & Milestones: (1) Connection of the GBOL infrastructure to NFDI-RDC; (2) API for taxonomic checklists for use at country level or even local level with FFH and Red List categories; (3) Extension of the GBOL-IT infrastructure for the inclusion of trait data; (4) Extension of GBOL portal and IT infrastructure for identification of organisms for e;g; Citizen Science projects.

5. Marine sample-based & geo-referenced organism biodiversity data

Contributors: HIFMB, AWI, BSH, UniBremen-MARUM

Challenge: Create a national open information system for marine sample-based geo-referenced organismic biodiversity data. The MARLIN (Marine Life Investigator) information network can become the backbone of a national information system for marine sample-based geo-referenced organismic biodiversity data, because (1) it is generic and scalable, i.e., it can cover the German EEZ in the Baltic, too, and it is open for expansion beyond the German EEZ, (2) its data model will be expanded to cover other data types (pictures) and other organism groups (3) MARLIN is open for data exchange with related and relevant platforms such as WORMS, Fishbase/Sealifebase and OBIS.

Current state: Since more than 10 years macro-benthos data from the North Sea EEZ are jointly managed. It took several years to turn data from numerous investigations into one coherent, quality-controlled data set. The information system MARLIN is based on a suitable data model, which fits the responsibilities of public administration, particularly regarding offshore windfarm impact on marine benthos, fish, mammals and birds. It will become operational in 2020.

Objectives & Milestones: (1) Secure taxonomic quality control: Develop and implement the interface between MARLIN and World register of marine species (WORMS) that enables automatic update of the taxonomic information in MARLIN; (2) Develop and implement structures and protocols that manage data ownership issues regarding raw data access, visibility and usability; (3) Establish the AWI/HIFMB-BSH data network, develop appropriate procedures and protocols that accommodate public administration restraints; (4) Expand MARLIN to cover plankton data, (5) Develop and implement procedures to track and secure the development of the MARLIN inventory; (6) Develop a suitable format for the data "collection" in NFDI4Biodiversity and develop and implement an interface between MARLIN and the "collection" level.

6. Land use/cover monitoring

Contributors: IÖR, UniBremen-MARUM

Challenge: Global and national land-use data (Copernicus Land monitoring services) are often insufficiently spatially and thematically resolved for detailed studies of biodiversity and its change. National official topographic geo data and OSM data are a better data basis here. However, such data must be prepared, processed, analysed, visualised and presented as a time series of indicator value development to enable long-term monitoring of land use/cover for biodiversity research.

Current state: The Leibniz Institute for Ecological Spatial Development develops in its research area "Monitoring" methods to collect land-use information small-scale and area-wide, to analyse it in combination with other data and to visualise it. The research data infrastructure "Monitoring of settlement and open space development (IOER-Monitor)" already contains a geodatabase with over 85 indicators that describe the land use situation on a small scale (100m grid data), in time series (from 2000 and annually from 2008) for the total area of Germany and make it available online (interactive geoviewer, WMS-, WCS- and WFS-Web Coverage Services).

Objectives & Milestones: (1) Compiling information on small-scale national land use/cover databases; (2) Contributions to the standardisation of land use classifications and nomenclatures; (3) Development of advanced indicators (including soil types, fertility, sealing, fragmentation, urban sprawl); Dissemination of information and training on research data on land use/cover.

7. Sample-based geo-referenced database, tools and services for metabarcoding analyses of freshwater ecosystems

Contributors: DNAquaNet, BGBM-Di, UniGießen

Challenge: Create an information system for geo-referenced freshwater metabarcoding samples that allows linking to different tools and services for enhanced biodiversity and ecological status assessments.

Current state: The importance of metabarcoding of freshwater samples is continuously rising as it allows for fast automated species identification across the whole freshwater biodiversity (microbes to fish). The data can be used to analyse ecosystem dynamics, responses to environmental drivers and interactions in much greater detail than traditional data. Up to now, no specialized data information system is available, even though the European and international community is already well connected and organized with over 500 participants from 50 countries actively contributing to COST Action DNAquaNet CA15219. Our objective is to enable stakeholders to access and analyse new metabarcoding data based on best-practice routines identified by the international DNAquaNet community.

Objectives & Milestones: Development of a suitable management system with taxonomic backbones und taxonomic quality control functionalities, that accommodates for data ownership issues. Priorities on macroinvertebrates, diatoms and fish. (1) Implementation of (semi-automated) tools and interfaces for cloud computing analyses based on best-practice analysis pathways identified within DNAquaNet; (2) Implementation of exploration/visualization tools.

8. Web portal "naturgucker.de"

Contributors: Naturgucker, UniGießen

Challenge: NFDI4BioDiversity will support naturgucker.de by the integration of services as well as IT developments in the sense and according to the needs of naturgucker.de and in strict adherence to the existing guidelines of the portal. This will be done with respect to visibility, use, teaching and playing opportunities in the field of organisms and to strengthen and support the observation of nature in Germany by volunteers.

Current state: The naturgucker.de portal is the largest information network for volunteer nature watchers in Germany. It currently includes more than 1.5 million images of nature objects of almost 38,000 species from more than 46,000 areas worldwide. In addition, this portal stores 10.5 million observations of plants, animals and fungi worldwide. More than 68,000 volunteer nature lovers participate in the social observation network naturgucker.de and commented with more than 570,000 participations observations and images of nature. Users are invited to report and document observations on various topics so that they can be evaluated together with co-operation partners. The project is supported by naturgucker.de with two strategic partners (federal state associations and Federal Association of NABU, see letter of support 11).

Objectives & Milestones: (1) Evaluation of collected data and attached metadata; (2) Optimizing visualization of data; (3) Development of cooperation interfaces; (4) Integration of taxonomic and / or geographical indices in the background; (5) Data mobilization, harmonization, aggregation and quality management; (6) Training of volunteers in the use and expansion of the naturgucker.de portal.

9. INSEKTEN SACHSEN Monitoring insects – Linking historical data with Citizen Science
Contributors: SGN, SNSB

Challenge: Creating interfaces and optimizing interoperability between different data infrastructures. Support in data standardization, implementation of ontologies and provision of access to NFDI4BioDiversity tools and services.

Current state: INSEKTEN SACHSEN [192] started in 2010 as a Citizen Science project. Citizen Scientists are recording insect observations, provided with photos for verification of species identification. Additionally, historical records from scientific collections, literature and diaries are edited and displayed in species-specific maps together with the recent observations from Citizen Scientists.

So far, there are more than 300,000 species-specific data records, together with over 65,000 photos and 2,000 pages. Challenges comprise the magnitude of data and their quality assurance, interoperability with databases containing data to the same topic, archiving and provisioning. Linkage to NFDI4BioDiversity services and tools can help to analyses population and ecosystem dynamics. Moreover, linkage with expert associations via NFDI4BioDiversity may help to enhance citizen scientists reports to INSEKTEN SACHSEN.

Objectives & Milestones: (1) Creating interfaces and optimizing interoperability between INSEKTEN SACHSEN data infrastructure and the NFDI4BioDiversity tools and services; (2) Provision of access to NFDI4BioDiversity tools, e.g. cloud computing analyses, exploration /visualization tools.

10. Developing an interface between NFDI4BioDiversity and NFDI4Agri

Contributors: Thünen-GDI, UniBremen-MARUM, IPK

Challenge: Creating interfaces and providing interoperability between the NFDI4Agri repositories reference implementation and NFDI4BioDiversity. Support in data harmonization between both NFDIs, implementation of transdisciplinary ontologies and provision of access to NFDI4BioDiversity tools.

Current state: To enable evidence-based assessment of biodiversity trends and factors causing these trends, the Thünen-Institute will develop and use a modular biodiversity monitoring program for agricultural landscapes in Germany. Data acquisition methods will range from nationwide general trend monitoring to in-depth monitoring of agricultural and regional focus areas and citizen-science based monitoring. Meanwhile, the Thünen-Institute is a member of the NFDI4Agri consortium, which will provide a reference implementation for a FAIR scientific data repository hosted on premise on institutional level. The biodiversity monitoring datasets will be stored in the reference implementation used at the Thünen-Institute.

Objectives & Milestones: (1) Creating interfaces and optimizing interoperability between the NFDI4Agri repository and the NFDI4BioDiversity tools and services, (2) Support with the implementation of standards and ontologies; (3) Provision of access to NFDI4BioDiversity tools, e.g. cloud computing analyses, exploration/visualization tools.

11. Linking regional **and national authorities to the NFDI4BioDiversity network and vice versa**

Contributors: LAU, UFZ, UniJena/iDiv

Challenge: Creating interfaces and optimizing interoperability between different data infrastructures. Support in data standardization, implementation of ontologies and provision of access to NFDI4BioDiversity tools.

Current state: Within the framework of international agreements, Germany has committed itself to the protection and sustainable use of biodiversity.

This also includes regular reporting of Germans biodiversity as 42% of all species in Germany are considered endangered, highly endangered or threatened with extinction. In order to halt the decline of biological diversity in Germany and to reverse this trend, knowledge about the condition and alteration of species and their habitats, as well as their threats, is of central importance.

Objectives & Milestones: (1) Creating interfaces and optimizing interoperability between the Scientific Monitoring Centre on Biodiversity and the NFDI4BioDiviversity tools and services; (2) Implementation of standards and ontologies; (3) Provision of access to NFDI4BioDiversity tools, e.g. cloud computing analyses, exploration/visualization tools.

12. Multitaxon data from a repeated sampling along an 1100 elevational gradient

Contributors: NP Bayerischer Wald, UFZ, UniJena/iDiv

Challenge: Create an information system for terrestrial geo-referenced multitaxon samplings that allows linking to different tools and services for enhanced biodiversity assessments.

Current state: The importance of multitaxon studies is continuously rising as it allows a broad link between biodiversity and the different ecosystem functions. The different response of different taxonomical and functional groups challenges the evaluation of climate change effects on biodiversity and thereby on ecosystem functions. NFDI4BioDiversity aims to support data management, analysis and computer capacity availability, such as routine analyses can be performed. Our objective is to enable administrations of protected areas to access and analyses their data as baseline for future management.

Objectives & Milestones: (1) Development of a suitable management system with taxonomic backbones und taxonomic quality control functionalities and that accommodates for data ownership issues; (2) Implementation of (semi-automated) tools and interfaces for cloud computing analyses; (3) Implementation for exploration/visualization tools.

13. Networking of protagonists in nature conservation

Contributors: HLNUG, UniGießen

Challenge: Networking of protagonists in nature conservation with the aim of harmonizing species references and field references as well as to continuously developing and maintaining them at a higher level. Establishment of a common species list of the cooperation partners with disclosure of the underlying concepts.

Current state: It takes a lot of time in nature conservation administrations to update and translate species lists of various origin. Information on reproduction, unit and collection method are quite different across different data sources. Even homonymous field references are defined differently within a species group. The federal states of Saxony (SMUL), Brandenburg (LfU), Mecklenburg-Vorpommern (LUNG), Saxony-Anhalt (LAU), Schleswig-Holstein (LLUR) and Berlin (SenUfK) are currently establishing a permanent cooperation within the "Ständiger Ausschuss Umweltinformationssysteme" (StA UIS).

A sub-goal is the harmonization of species and field references between these states. Data exchange with the platforms NATURGUCKER and ORNITHO has already begun and will be expanded.

Objectives & Milestones: (1) Development of tools to match species references and ensure permanent support and development of tools; (2) Creation of an ontology for referencing species found in the field; (3) Documentation of field reference ontology; (4) Provision of simple tools to transform field references and secure support.

14. Enabling the AlgaTerra information system for the connection to tools and services for e.g. web-/cloud-based metabarcoding analyses

Contributors: BGBM-DI, BGBM, UniGießen

Challenge: Develop services and tools for subsequent metabarcoding analysis to facilitate enhanced biodiversity assessments. Provide sequences and correlated sequence information (e.g. ecological, taxonomical metadata) coming from a taxonomically curated database.

Current state: The AlgaTerra information system is offering evaluated information on synonyms and concepts. In addition, it is providing taxonomic, molecular and ecological data of mainly non-marine microalgae. This data is based on specimens, samples and cultures deposited in Natural History Collections and enables researchers to perform e.g. a web-based BLAST search on vouchered sequences. The focus of AlgaTerra is on diatom specimen collections as well as documented living cultures linked to collection- and DNA bank specimens.

Objectives & Milestones: (1) Development and implementation of a web service to enable automated BLAST search for cloud computing analyses which provided number of matches, number of mismatches, reason, scientific names from the curated taxonomic database, strains, specimens, INSDC numbers, links to species pages in AlgaTerra including morphological and environmental information.

15. Connecting the natural history societies to the NFDI4BioDiversity network

Contributors: GdO-Libellen, UniJena/iDiv

Challenge: Creating interfaces and optimizing interoperability between different data infrastructures. Support in data standardization, implementation of ontologies and provision of access to NFDI4BioDiversity tools and services.

Current state: More than 1.5 million species-specific data records are being hosted by the odonatologists in Germany. Scientific analyses require data with high spatial, temporal, taxonomic and functional resolution to understand how environmental drivers affect ecosystem functioning, goods and services. The scattered data landscape needs re-evaluation of data granularity and cross-linkage to climatic, soil or other additional data. Moreover, linkage with expert associations via NFDI4BioDiversity will help to enhance citizen scientists reports to the database, e.g. monitoring of around 80 dragonfly species and their populations.

Objectives & Milestones: (1) Creating interfaces and optimizing interoperability between the different data infrastructure on the federal level and the NFDI4BioDiversity tools and services; (2) Provision of access to NFDI4BioDiversity tools, e.g. cloud computing analyses, exploration/ visualization tools.

16. Connecting the data of the natural history society AraGes e.V. to the NFDI4BioDiversity network with a German monitoring portal and vice versa

Contributors: AraGes-Spinnen, SNSB, SMNS, UniJena/iDiv, Staatliches Museum für Naturkunde Karlsruhe, RWTH Aachen

Challenges: Enhancing standardization and thus optimizing interoperability of the ARAMOB and ATLAS data through new interfaces. Implement taxon relevant ecological lists, thesauri and ontologies and provision of access to NFDI4BioDiversity tools and services.

Current state: More than 1 million species-specific data records are being hosted by the ATLAS DER SPINNENTIERE EUROPAS of Arachnologische Gesellschaft and more than 70.000 species-specific data records by the arachnological database ARAMOB. Linkage to the NFDI4BioDiversity services and tools can help to analyses ecosystem dynamics. The scattered data landscape needs re-evaluation of data granularity and cross-linkage to climatic, soil and additional data. Moreover, linkage with expert associations via NFDI4BioDiversity may help to enhance citizen scientists' reports to the Atlas database, e.g. monitoring of spider taxa, species, species groups, assemblages and others.

Objectives & Milestones: (1) Standardization of AraGes data products for interoperability with NFDI-RDC and the German monitoring portal; (2) Implementation of web interfaces and APIs to optimize interoperability a) between the AraGes IT infrastructure, the "ecological database ARAMOB" and the "record database of the ATLAS", and NFDI4BioDiversity tools and services; (3) Development and enhancement of NFDI4BioDiversity accessible thesauri for habitat conditions relevant for spiders (Araneae) with DWB management tools (4) Improvement for member management, data influx and graphic user interface.

17. Netzwerk Phytodiversität Deutschlands e. V. (NetPhyD): Connecting natural history societies to the NFDI4BioDiversity network and vice versa

Contributors: NetPhyD-Pflanzen, UniJena/iDiv

Challenge: Creating interfaces and optimizing interoperability between different data infrastructures. Support in data standardization, implementation of ontologies and provision of access to NFDI4BioDiversity tools and services.

Current state: More than 30 million species-specific data records are being hosted by the Deutschlandflora portal. Challenges comprise data collections, quality assurance, archiving and provisioning. However, in order to monitor changes and halt the decline of biological diversity in Germany, knowledge about the condition and alteration of species and their habitats, as well as their threats, is of central importance.

Linkage to the NFDI4BioDiversity services and tools can help to analyse ecosystem dynamics. Moreover, linkage with expert associations via NFDI4BioDiversity will help to enhance citizen scientists reports to the Deutschlandflora database, e.g. monitoring of moderately frequent taxa as those are candidates for red lists with the least current knowledge of trends.

Objectives & Milestones: (1) Creating interfaces and optimizing interoperability between the Deutschlandflora data infrastructure and the NFDI4BioDiversity tools and services; (2) Provision of access to NFDI4BioDiversity tools, e.g. cloud computing analyses, exploration/visualization tools; (3) Communication and regular information exchange for mutual benefits.

18. Connecting German Ichthyological Society to the NFDI4BioDiversity network and vice versa

Contributors: GfI-Fische, UniJena/iDiv

Challenge: Creating interfaces and optimizing interoperability between different data infrastructures. Support in data standardization, implementation of ontologies and provision of access to NFDI4BioDiversity tools and services.

Current state: In the "Digital Fish Species Atlas of Germany and Austria" more than 100,000 data sets of fish species have been collected since 2003 by GfI (German Ichthyological Society). The data comes from literature searches, database queries and increasingly also from Citizen Science. The Atlas is unique for Germany and Austria, representing all regional freshwater and marine fish species. The distribution maps in connection with species descriptions are managed with the software "Biodiversity Warehouse". It prepares data for internet presentation and is also suitable for the building of other faunistic distribution atlases (e.g. mussels and mammals of Bremen, fish of Pernambuco/Brazil).

The challenge in maintenance and further development of the database includes the collection of data, the provision of interfaces for export to users from science, administration and environmental education or the transfer of data to global databases. By linking it to climate, soil or other additional data via NFDI4BioDiv, a significant added value in data use can now be achieved.

Objectives & Milestones: (1) Creating interfaces and optimizing interoperability between the "Fish Atlas of Germany and Austria" and NFDI4BioDiversität tools and services; (2) Provision of access to NFDI4BioDiversity tools, e.g. cloud computing analyses, exploration/visualization/data exchange tools; (3) Communication and regular information exchange for mutual benefits.

19. Geo-Visualization tools for biodiversity data

Contributors: IfL, UniJena/iDiv

Challenge: Spatial visualizations with maps, are indispensable for the communication and effective dissemination of biodiversity data. Using maps, large amounts of data can be made easily accessible and analysed.

But it is precisely the visualization of large and inconsistent amounts of data that poses challenges. Maps that are too complex or have content that is difficult to interpret can also have a deterrent effect on users.

Current state: The aim is to develop suitable visualizations for different user groups that consider the specific characteristics, individual data sets quality, their spatial characteristics and existing conservation interests. The visualizations should also support the interpretation of the data, e.g. by combining relevant data sets or by providing information on the reliability of the surveys. These developments are to be based on experience gained in the BMBF-funded Citizen Science project "Experiencing species diversity: How nature research benefits from interactive web maps on your own doorstep" (2018-2020). In this case, ornithological observation data were visualized in a web map application, for example in combination with information on observer density.

Objectives & milestones: Develop and test visualization methods and implement individual methods as tools for interactive web map presentations. (1) Capturing the wishes of different user groups for visualizations, (2) ratio of occurrence/observations to total number of observations; (3) Relationship of different species/species groups to each other, e.g. combination of bird observation data with those of insects, in order to be able to draw conclusions about changes in the population; (4) Consideration of different conservation statuses of species and different spatial and/or temporal resolutions of the collected data; (5) Presentation of one's own observations as motivation for further surveys and in order to be able to recognize one's own contribution.

20. Research server for national parks

Contributors: NP Hunsrück-Hochwald, UFZ

Challenge: Establish a uniform, comparable and interdisciplinary documentation of the national parks research projects, with the aim of making them more accessible to the (specialist) public and finally to the NFDI. This basic feature is missing for most national parks in Germany.

Current state: Actually, the metadata information system with three parallel available catalogues and the integration of a literature database, as well as specific websites is in operation. The connection of a geoserver and geoweb services is realized.

Strategic view: The national parks are collecting lots of data, some already for several decades. However, the accessibility to (meta-) information about these data is poor. The embedding of the research server in NFDI4BioDiversity will make it even more attractive for other national parks, creating an opportunity to represent the data of more national parks in the future.

Objectives & Milestones: (1) Extension of the information platform with components of a document management system; (2) Implementation of appropriate interfaces to NFDI; (3) Connection or coupling with systems such as ALA; (4) Attracting other national parks to join the project and feed in their data.

21. Networking of stakeholders in nature conservation, improving the quality and readability of databases

Contributors: LfULG, UniJena/iDiv

Challenge: Establishment of minimum technical and professional standards for the collection and documentation of species observation data. Development of tools for the consolidation of existing databases. The concepts and tools must be simple and limited to essential factors, facilitating the use for different actors in nature conservation.

Current state: Since 2008, a central species database exists in Saxony, which serves as a state collection, documentation and information system for fauna and flora. Meanwhile, the state authorities of Brandenburg (LfU), Mecklenburg-Vorpommern (LUNG), Saxony-Anhalt (LAU), Hesse (HLNUG), Schleswig-Holstein (LLUR) and Berlin (SenUfK) are applying the same software (MultiBaseCS) and set up a working group. Among other things, this involves the networking of the state authorities with the aim of establishing minimum standards for the collection and documentation of species observation data, thereby improving the quality and readability of data.

Objectives & Milestones: (1) Information exchange and Integration of already existing projects with overlapping content; (2) Providing efficient tools for the delivery of species observation data (INSPIRE, Red List Germany, etc.) to the Federal Agency for Nature Conservation (BfN) or to EU; (3) Nationwide clarification of legal questions regarding the handling of species observation data (UIG, data protection); (4) Ensuring continuity of continuous tasks (e.g. maintenance of species reference lists); (5) Agreement on nationwide minimum standards, which are aligned with existing IT solutions of federal states; (6) Nationwide permanently maintained species reference list; (7) Provision of tools for matching type references; (8) Establishment of species-specific minimum technical standards for the collection and documentation of species observation data.

22. Leveraging NFDI for the iDiv community – Leveraging iDiv Data for the NFDI community

Contributors: MPI-Biogeochem, UniLeipzig, UniJena/iDiv

Challenge: For many research questions addressed by the German Center for Integrative Biodiversity Research (iDiv) community, NFDI4Biodiversity curated data will only be interpretable with those embedded in the NFDI4Earth. This use case aims to showcase how to leverage the joint potentials of the emerging NFDIs.

Current state: Today, it requires considerable manual effort to integrate the separated curated earth system and biodiversity data. Over the past five years, iDiv has become a key nucleus for innovative data sets in biodiversity research. Within iDiv, the PlantHub initiative was recently initiated to provide a unified entry point to the wealth of plant-related data hosted by iDiv including, for example, functional plant traits TRY (>11.8 Mio entries for 280,000 plant species), global vegetation plot data sPlot (1.1 Mio plots) and the Leipzig Catalogue of Plants. These data shall be used jointly to address fundamental ecological questions but also to support efforts to monitor biodiversity and ecosystem change in the framework of GEO BON.

DFG

To do so, these data sets need to be combined with data sets at varying spatiotemporal resolutions. Information of this kind will be provided in the future via the NFDI4Earth likely using "Earth System Data Cube".

Objectives & Milestones: (1) Semantic integration of PlantHub data sources (2) Integration in the NFDI-RDC; (3) Development of APIs for data sources from NFDI4Earth for integration in the NFDI-RDC; (4) Development of data stories showcasing the benefit of cross-NFDI data integration.

23. Thematic network on biodiversity data to scope ALA instance

Contributors: UniJena/iDiv, BGBM

Challenge: Establish a thematic network on biodiversity data with the German user community to develop a uniform, interdisciplinary and interoperable documentation platform of community-driven science data including relevant agency data, to enhance their visibility, interoperability and visualisation, making them more accessible to both interested experts and the public. In close collaboration with the data collectors mentioned above, in particular those involved in our other use cases, we will design and set up a platform (based on the ALA software) to develop a highly functional, attractive and powerful background integration infrastructure for users to make their data interoperable and gain enhanced visibility through linkage. This will happen with a 3-phase work plan moving from mobilisation of selected data to full integration into the NFDI-RDC.

Current state: Currently many organisations (natural history societies, NGOs, citizen science projects, national parks and agencies) are collecting valuable data, often already for decades, but the accessibility to and visibility of (meta-) information of these data is poor.

Objectives & Milestones: (1) Development of thematic network with annual user conferences and four expert workshops per annum; (2) Implementation of a customizable data platform in close collaboration with the user community and GBIF, incl. available tools sets; (3) implementation of appropriate interfaces to the NFDI-RDC (TA4 M4); (4) Attracting further data providers to join and link their data.

Milestones:

Milestones	Description	Month
M1.1.1	Establishment of a science-society-policy advisory board with community representatives	4
M1.1.2	Start of thematic network with annual series of user community conference and 4 expert workshops series focusing on the needs of different communities (authorities, NGO's, natural history societies, research institutions,), in collaboration with use cases	9
M1.1.3	Ticket system for data support and data protection advice agreed and documented, established	24
M1.1.4	NFDI4BioDiversity strategy for use case support agreed and documented	36
M1.1.5	Field report and analysis published in peer reviewed journal	60

TA1 M2: Support with tools/platforms for (early) mobilization of data and data analyses

Lead: GFBio e.V.; **co-lead:** UniJena/iDiv, HITS

Contributors: DSMZ, SNSB, UniGießen, BIBI, FBN, GWDG, HIFMB, ZM, IGB, Naturgucker, SGN, BGBM-DI, AraGes, DDA-Vögel

Description: NFDI4BioDiversity recommends and supports the application of easy-to-use and stable biodiversity data tools and services for data mobilization and analysis to facilitate entering the data life-cycle at the earliest possibility. Currently, many communities rely on spreadsheet tools for data documentation, as these are easy to use. Even if tools like Rightfield [106] allow semantic annotation in spreadsheets, data quality and interoperability with other systems remain a major challenge. The highly valuable biodiversity inventory and monitoring projects done by citizen science communities (e.g. learned societies) often have domain-specific data portals and Apps, where guided data entry and data access is cooperatively performed. However, quality and plausibility assurance, taxonomic backbones and interoperability of data is often lacking. To effectively improve the situation, NFDI4BioDiversity will support the data providers by 1) recommending tools, which are missing in their repertoire or better fulfil common standards in data acquisition, 2) custom-tailored training and IT-support in tool development, deployment and use, and most importantly 3) direct, hands-on support in adapting their own tools and platforms towards interoperability with the NFDI-RDC. Hereby, NFDI4BioDiversity will address a wide array of community demands in 1) gathering, structuring and managing raw data, 2) realizing scientific workflows in the field and lab, 3) quality and plausibility assurance of data, 4) analysing and modelling biodiversity data, 5) creating joint virtual research environments for networking and sharing project data 6) publishing data, and 7) archiving biodiversity data objects.

This measure will provide direct technical support to users with tools to structure, validate and standardise their data, ensure its interoperability and ultimately, improve its FAIRness. Together with SIG2 we will advise on the adaptation and deployment of tools and platforms for gathering, structuring and (early) mobilization of data, and tools for project data management and analyses. We will focus on widely used, preferably generic tools, such as the map and geo-portal solutions used by the LTER-D platform [55], the GFBio workbenches for early data management and transformation [235] and the ALA Software [234] in the ALA use case (TA1 M1). We will employ methods of co-design and co-production with the data providers and community software developers, integrating our input directly into their work process in an agile manner. Together with SIG3 we will recommend appropriate training modules, provided by TA1 M4, and guide the users through applying their newly acquired skills. In cooperation with TA3 M1-2 and TA4, we will ensure the tools and platforms we recommend and support comply with the concept and technical requirements of the NFDI-RDC, so that users can profit by deploying their software on RDC-approved platforms or exchange data via standardized APIs.

To leverage the great technical diversity of the biodiversity software ecosystem, we will extend our network of expertise in bioinformatics, ecoinformatics, biodiversity informatics and general data management.

Milestones:

Milestones	Description	Month
M1.2.1	Technical support with deployment of tools and platforms started, support by GFBio workbenches started	12
M1.2.2	Support workflows and knowledge sharing with TA3 M1-M2 established.	24
M1.2.3	NFDI4BioDiversity strategy for RDM tool support in accordance with NFDI-RDC for science projects agreed and published	36
M1.2.4	NFDI4BioDiversity strategy for RDM tool and platform support in accordance with NFDI-RDC for independent research data web portals and citizen science monitoring projects agreed and published	48
M1.2.5	Field report and analysis published in peer-reviewed journal	60

TA1 M3: Front office/back office: Active Research Data Management (RDM) & data science support

Lead: GFBio e.V.; **co-lead:** SNSB

Contributors: BGBM, DSMZ, MfN, SGN, SMNS, SNSB, ZFMK, UniBremen-MARUM, AWI, UniJena/iDiv, IPK, HITS, GFBio e.V., SUB, ZMT, AraGes, Naturgucker, DNAquaNet

Description: The RDM strategy of NFDI4BioDiversity relies on a close cooperation with the providers and users of research data throughout all stages of the data life-cycle. The work of GFBio has revealed a number of general RDM challenges related to guidelines and legal regulations for environmental and biology data and material (e.g. Convention on Biological Diversity [236], Intellectual Property Rights and licenses) as well as community-agreed content schemas and standards for mobilising, exchanging and archiving biodiversity (meta-)data. In this measure, we will offer personalized expert support on these topics to individual researchers and research groups. For this, we will extend our successful GFBio front-office/back-office helpdesk model and ticket system. The front-office will be responsible for the direct communication with the data providers and users, preferably "on-site" at the universities and institutes. This will be realised by significantly expanding our already established network of data managers. In this case, the NFDI4BioDiversity helpdesk team will act as the back-office. We will also take care of support requests addressed directly to our helpdesk team. Together we will provide access to experts on general RDM and manual curation, validation and transformation of scientific data from the various domains of taxonomic research, ecology, environmental science, and (molecular) biodiversity research. We will 1) guide researchers through the process of publishing well-structured and re-usable data products, 2) advise on creating data, (re-)structuring data, data quality and enrichment by using community-agreed vocabularies and ontologies,

3) support the application of major standards for biodiversity data (e.g. ABCD, DwC, EML, MIxS, SDD, MOD-CO), 4) support the curation and publication of terminologies and the long-term provision of scientific software, 5) support biodiversity researchers and volunteer experts with the development of data management plans. The helpdesk team will have access to the expertise of the whole consortium to provide the best possible service while maintaining a single point of contact for the front-office and users/providers. Specifically, but not exclusively, we will be able to offer support for 1) biodiversity software developers via TA3 M2, 2) long-term provision of scientific software (incl. descriptive metadata), compatible with NFDI-RDC via TA4, 3) tools offered by TA1 M2, 4) relevant national and international services reviewed by TA2 M2 and TA3 M1 and 5) NFDI-RDC recommended tools and services via the "test and certification help desk" described under TA3 M2. The scope of the support will be continuously extended to cover fast-evolving biodiversity topic areas such as AI, 3D image analysis, streaming data and lab workflows.

Milestones

Milestones	Description	Month
M1.3.1	Back-office operational, ticket system unified and consortium trained	12
M1.3.2	Training workshops for front-office personnel at universities and institutes	24, 36, 48, 60
M1.3.3	Full integration with certification helpdesk (TA3 M2) established	24
M1.3.4	Front-office model established at several additional locations	48, 60

TA1 M4: Education and training

Lead: UniMarburg; **co-lead:** UniGöttingen

Contributors: UniJena/iDiv, GWDG, UFZ, IPK, BIBI, SUB, GFBio e.V., GdO-Libellen, DDA-Vögel, GfÖ, all universities

Description: The goal of our education activities is to establish the teaching of FAIR research data management according to the FAIR4S-framework of the EOSC [237] as a core topic in the undergraduate and graduate curricula of biodiversity-related degrees at German universities and to thus ensure that current and future generations of scientists will possess profound knowledge in these topics. To reach this goal UniMarburg and UniGöttingen will establish an education and training platform imparting research data management and data literacy skills (see [238]). Coordinated by the training coordinator (SIG3) UniMarburg and UniGöttingen will offer training on all phases of the data life-cycle with a special focus on the provisioning, management and use of high-quality FAIR data. The platform will contain curated datasets and case studies covering different aspects of the data life-cycle, closely linked to the NFDI4BioDiversity use cases (TA1 M1) and objectives. The platform will be available for researchers, citizen scientists and students to obtain hands-on knowledge on the services and tools provided by NFDI4BioDiversity and to sharpen their RDM skills.

Co-designed with our university partners it has been agreed that the participating universities will integrate teaching materials offered by NFDI4BioDiversity and the NFDI into their curriculum. The newly established data literacy curriculum at UniGöttingen, the training offers at the Marburg Service center for eResearch, the DATAx project at Leuphana University (see letter of support 10) and the RDM helpdesk at UniJena/iDiv will be our pilots. On the graduate level, an important partner will be the cross-domain Bremen graduate education program on research data management and data science (see "Graduate Education" in Section 2.3) and the Göttingen Graduate Center for Neurosciences, Biophysics, and Molecular Biosciences (GGNB) [239]. These activities will be closely coordinated with other NFDI consortia as well as with central NFDI data and training resources by SIG3. The target group of these training is essential for spreading usage of the NFDI, as graduate students lie at the heart of data collection, transformation and usage.

Raising awareness of NFDI4BioDiversity, its services and training opportunities among multiplicators is another goal of this measure. With the help of the SIG1 coordinator and our participants (GdO-Libellen, DDA-Vögel, GfÖ) we will target VBIO (see letter of support 15) [240], GfBS [241], DZG [242], NEFO [243], Botanical Societies, Zoological Societies and their respective communities, universities and biodiversity research infrastructures and aim to reach them by attending meetings and dedicated visits and via social media and the provisioning of training videos.

NFDI4BioDiversity will strongly work towards connecting the consortium-specific platform and materials with other NFDI-consortia and with overarching NFDI-services in education and training as part of our engagement in cross-cutting topics.

Milestones:

Milestones	Description	Month
M1.4.1	Creation of training datasets and teaching videos	12
M1.4.2	Design of case studies in collaboration with other NFDIs and TA1 M1 use cases	24
M1.4.3	First case studies are incorporated into university teaching	36
M1.4.4	Outreach to multiplicator communities established	48
M1.4.5	Revision of case studies based on first practical experiences	60

TA1 M5 Regular quality surveys, user feedback, quality management, reputation systems

Lead: GFBio e.V.; **co-lead:** UFZ

Contributors: UniJena/iDiv, UniMarburg, UniGießen, BIBI, BGBM, FBN, naturgucker, UniGöttingen, GdÖ-Libellen, DDA-Vögel, GfÖ, the use case community (TA1 M1)

Description: The RDM strategy of NFDI4BioDiversity relies on a close cooperation with the providers and users of research data. Collecting, evaluating and implementing user feedback is a key instrument to successfully serving the biodiversity community. To this end, GFBio e.V. and BIBI together with SIG2 will design and conduct different surveys to monitor:

- users' satisfaction with the use case implementations (TA1 M1) and with support on tools for data mobilization (TA1 M2), RDM user helpdesk (TA1 M3) and certification and testing helpdesk (TA3 M2); continuously
- the effectiveness of outreach activities (TA1 M4); continuously
- the quality of NFDI-RDC services (TA4 M5); continuously
- evolving community demands; annually
- cultural changes in newly developed collaborative environments, trends in data use and reuse, the impact of FAIR data on the reward system(s) in biodiversity research; biennially

The results will be used to iteratively improve the quality and effectiveness of our measures. Together with TA5 M1, we will implement an overarching NFDI4BioDiversity quality management strategy, aligned with the ISO 9001 standard for quality management systems. Quality management will ensure that any kind of user feedback, is used for improving the tools and services we offer in a structured and effective way, ultimately increasing their acceptance by the community.

One of the main pillars of FAIR data is the initiated cultural change, which deals with the incentives for data producers to invest the extra effort in FAIRifying their data [244]. Since cultural change, reputation and credit systems have been defined as a central cross-cutting topic, in this measure we are committed to contributing our ideas to value the work of data producers for contributing FAIR data and software.

Milestones:

Milestones	Description	Month
M1.5.1	Instant surveys for user satisfaction implemented (based on preliminary work in GFBio)	6
M1.5.2	Service quality management strategy implemented	12
M1.5.3	Surveys for community demands and cultural change implemented	18
M1.5.4	Common concept of credit and reputation system discussed	36
M1.5.5	Results of survey activities published	54

DFG

4.3   Task Area 2: National and international networking (2connect)

Lead: BGBM; **co-lead:** GWDG, BIBI

Aims & overview: NFDI4BioDiversity is born into an evolving landscape of international initiatives and infrastructures. Due to the excellent networking already in place (e.g. with/in GBIF, GEO BON, CETAF, DiSSCo, ELIXIR, de.NBI, iBOL and other NFDIs) and, in many cases, through the coordinating functions of NFDI4BioDiversity actors, the consortium is ideally equipped to create effective cooperation with these initiatives. This task area will 1) coordinate (inter)national implementations, 2) identify development priorities, and 3) harmonise access to existing services by applying standards at all relevant levels. The respective measures will be flanked by use cases and user stories (through TA1) underpinning different aspects of the use of standards for the interoperability of data and services both within the community and across disciplines. Last, but not least, user stories will also help to calibrate draft standards prior to the respective reviews and ratifications.

TA2 M1: Collaboration with related NFDI consortia

Lead: BGBM; **co-lead:** GWDG

Contributors: All co-applicants and participants with their networks

Description: NFDI4BioDiversity is committed to take over an active role in overarching working groups within the NFDI landscape. As the respective structures are yet to be created, NFDI4BioDiversity strongly advocates together with 21 other NFDI consortia to do this according to the "Berlin Declaration on NFDI Cross-Cutting Topics" [22]. Independent of the actual implementation, we will participate in the foreseen joint activities to identify common requirements for core services across different NFDI consortia in cooperation with SIG5. To this end, we will derive a set of requirements for horizontal services within the NFDI. For instance, federated identity management, authentication and authorization infrastructure, persistent identification, federation of repositories and (standardized) metadata services are seen as natural contenders here. NFDI4BioDiversity partners have already established such services, which could be made available to other consortia, and is committed to align its infrastructure towards common solutions. Due to the cross-disciplinary nature of NFDI4BioDiversity data, we also see an urgent need to continue our work on common data models, their interoperability and standardisation (addressed in NFDI4BioDiversity in TA4). Part of this discussion will take place within and across the German NFDI consortia and NFDI4BioDiversity is committed to contribute its expertise.

Contributions to NFDI working groups: BGBM and GWDG will compile requirements for common services that can be used to align requirements across NFDI consortia.

NFDI4BioDiversity is ready to engage and to lead such endeavours, and it offers to realise reference implementations for cross-consortia use of services.

DFG

As NFDI4BioDiversity builds upon the existing infrastructure of GFBio, we can offer our current service-based architecture (as described in Section TA4) as a starting point for establishing interoperability and federation. Where necessary and sensible, we will offer the development, provisioning and maintenance of common services to other consortia, best as part of the NFDI-RDC. Similarly, we plan to reuse solutions from other NFDI consortia to prevent duplication of work.

Providing common use-cases for interoperability: To facilitate the concept of a ONE NFDI with common services and a high degree of interoperability, we propose the identification and implementation of common use-cases. The respective partners in TA1 will actively engage in this process by contributing ideas for such use-cases, specify them in collaboration with other consortia, as well as implementing them to demonstrate the practical advantages of a common NFDI. To prime this process, cross-domain use cases with NFDI4Agri and NFDI4Earth (see TA1 M1 use case 2, 10 & 22) have already been filed.

This measure is closely interlinked with the requirement analysis of TA1 and TA3. The further development is expected to be a cross-cutting coordination between all NFDI consortia to be further specified in the general NFDI organization. As the NFDI-RDC will need to be synchronized with the international landscape, the measure is also dependent on TA2 M2.

Milestones

Milestones	Description	Month
M2.1.1	Common requirements for core services across NFDI consortia identified	12
M2.1.2	Providing common use-cases for interoperability	24
M2.1.3	Common data models identified and standardized	36
M2.1.4	Requirements for common services across consortia filed	48
M2.1.5	Implementation of cross-consortia services within NFDI-RDC finished	60

TA2 M2: Context with international services & developments

Lead: BGBM; **co-lead:** UFZ

Contributors: BIBI, DDA-Vögel, DNAquaNet, DSMZ, FBN, GFBio e.V., GWDG, HITS, IPK, Naturgucker, SGN, SUB, Thünen-GDI, UniGießen, UniJena/iDiv, UniMarburg

Description: NFDI4BioDiversity will be embedded in an evolving landscape of international initiatives, projects and services, as the challenges of rapidly evolving data-driven international biodiversity research can only be met in collaboration with strong partners. This measure will ensure that NFDI4BioDiversity avoids the implementation of parallel structures and instead establishes itself with a clearly defined portfolio of powerful services as part of the global infrastructure and benefits taking advantages from existing offerings. Based on continuously expanding liaisons with relevant partner initiatives, we will be particularly active in three areas:

Contributing to international working groups: The essential developments of the community like data formats, protocols, APIs, services, or synchronization strategies are prepared and specified in international working groups. It is therefore essential to contribute to these developments at the earliest possible stage in order to ensure that NFDI4BioDiversity issues are taken into account and to incorporate international developments into one's own infrastructure planning. Consequently, NFDI4BioDiversity partners have already coordinated and will further enforce their participation in working groups under the umbrella of RDA, the EOSC Secretariat, DiSSCo, CETAF and GBIF, BGBM and UFZ will support all contributing partners in documenting, analysing and strategically expanding their efforts to ensure that all relevant working groups are decently covered.

Service coordination: NFDI4BioDiversity is designed as a service infrastructure with platform-independent and machine-executable core functions provided to interact in powerful research workflows. Availability, performance and robustness of the underlying basic services are of essential importance for the long-term sustainability of this approach. All contributors will, therefore, analyse the required services of external infrastructures in their domain of expertise (e.g. GBIF [3], EUDAT [18], DataONE [19], Catalogue of Life [245]) and formalize reliable access to these services with the respective cooperation partners. A particular focus will be on the assurance of the stability of the service APIs integrated into NFDI4BioDiversity workflows coordinated by SIG2. The relevant basic services will be documented and described in an open registration system (see TA3). NFDI4BioDiversity partners (BGBM and SGN) are involved in an initiative within the framework of TDWG Biodiversity Information Standards (partners BGBM, IPK and SGN) [7] and DiSSCo (partners BGBM and SGN, see letter of support 5) [191] to re-implement the BiodiversityCatalogue registration system. NFDI4BioDiversity partners (BGBM and SGN) are involved in an initiative within the framework of TDWG Biodiversity Information Standards [7] and DiSSCo [191] to re-implement the BiodiversityCatalogue registration system. In this context, it will be evaluated whether NFDI4BioDiversity can operate the system and make it available to the international community.

Data Use Agreements: Despite the international movement to make research data freely and openly available, there is still a large number of data sources, representing an important basis for answering societal relevant research questions, that is not freely available or free but not accessible. NFDI4BioDiversity will organise through its partners BGBM and UFZ access to urgently needed data sources within the framework of Data Use Agreements (DUA). In many cases, these data have been collected outside biodiversity research (e.g. data on transport infrastructures, which are of great importance for the analysis of biodiversity patterns and their changes). In addition to mobilizing the data for research, the aim is to provide standardized DUAs that can be used as a blueprint for fast access to previously inaccessible data.

DFG

Formal agreements on the mutual use of services and datasets make an important contribution to the sustainability of the NFDI4BioDiversity infrastructure by avoiding double implementation and reducing costs.

Milestones:

Milestones	Description	Month
M2.2.1	First strategic report on international involvement completed	12
M2.2.2	NFDI4BioDiversity Open Registration System up and running and NFDI4BioDiversity services registered	24
M2.2.3	2nd strategic report on international involvement completed	36
M2.2.4	NFDI4BioDiversity Open Registration System fully operational and accessible via API	48
M2.2.5	Data/service Use Agreements (DUAs) signed	60

TA2 M3: Engagement in European and international standardisation activities

Lead: UniBremen-MARUM; **co-lead:** AWI

Contributors: BGBM, SNSB, UniJena/iDiv, IPK, DSMZ, HITS, FBN

Description: For NFDI4BioDiversity, the use and support of standards is an indispensable basis for the interaction of infrastructure components as well as the sustainable provision of data for future interdisciplinary research activities. The standards used refer mainly to biodiversity data, metadata and semantic information, but also deal with more generic access protocols, workflows and best practices for, e.g., FAIR practice, persistent identifiers, data annotation and cross-linking (for details please refer to section 3.1).

Many NFDI4BioDiversity consortium partners are already involved in standardisation committees, often in a coordinating capacity. As part of this measure, AWI and UniBremen-MARUM will extend and structure these activities to ensure that the standards required for the NFDI-RDC infrastructure (as specified in TA4) are available. In the first phase, a landscape analysis will be carried out to establish a structured list of the standards and best practices required for NFDI4BioDiversity as well as the remaining gaps and the standardisation bodies and initiatives required to close them (e.g. GSC, TDWG, W3C, CETAF, GGBN, etc.). In the second phase, standards and best practices will be prepared in cooperation with these committees and tested and refined within the framework of NFDI workflows. The submission, ratification and publication of standards are planned for the final phase. Community-independent standards (e.g. for web annotation and linked open data) will be discussed in cross-NFDI working groups in order to exploit the potential of a common approach to standardisation and the use of standards.

Milestones:

Milestones	Description	Month
M2.3.1	Standards landscape analysis draft circulated	12
M2.3.2	Standards landscape analysis published in peer reviewed journal	24
M2.3.3	Standards assessment and calibration documented on NFDI4BioDiversity communication platform and shared with other NFDI consortia	42
M2.3.4	Ratification process started	42
M2.3.5	Ratification completed and standards documented	60

4.4 Task Area 3: Long-term data, tools and service preservation, certification (2consolidate)

Lead: UniJena/iDiv; **co-lead:** SNSB

This task area consolidates existing German infrastructures and data pipelines for long-term FAIR data preservation and publication. Such pipelines have been established in the last years, e.g. in the context of GFBio and related national and institutional infrastructures and portals. Furthermore, this task area fosters consolidation and enhancement of existing services and tools given they are open source, under active development and well-accepted by the biodiversity user community and/or by domain-specific data archives or biodiversity data publishers in Germany. This concerns a number of discipline-specific services, which are in place at data repositories, publishers and data brokers (e.g. GFBio with its approved data centers [246]), and tools for handling bio-data already in the early stage of the data life-cycle (e.g. BEXIS2, DWB and others offered by de.NBI and bio-tools [233]). The basic operation of services and tools belonging to the NFDI4BioDiversity consortium is provided by the NFDI co-applicants as in-kind contributions. The provision and technical enhancement of NFDI4BioDiversity-adapted GFBio core services, data pipelines and workflows for direct support of data publication efforts are subject of TA3 M3, their connection and integration in the NFDI-RDC is done in cooperation with TA4 M1. In this task area, we screen and document the current and emerging service and tool ecosystem in biodiversity research. It scouts for emerging research topics and opportunities, as well as new types and formats of data and workflows, infrastructure, and service requirements in the biodiversity domain. The survey and evaluation of evolving new data management, data analysis and data archiving strategies in the wider biodiversity community landscape in TA3 M1 will be complemented by documentation and consolidation activities in TA3 M2 and M3. The focus will be on tools for data management early in the life-cycle and FAIRness (together with TA1 M2, SIG2) and software for data validation, data transformation and data quality assurance. As soon as data centers, services and tools are recommended by NFDI4BioDiversity they are profiled and guided towards certification (TA3 M2). For certification issues, this task area will work with NFDI4RSE and other cross-cutting structures.

TA3 M1: Identification and description of the current and emerging technical ecosystem in biodiversity research

Lead: GWDG; **co-lead:** IPK

Contributors: SNSB, BGBM, UniBremen-MARUM, GFBio e.V., BIBI, HITS, UniGießen, SGN

Description: This measure will consolidate and adapt existing work of GFBio and de.NBI, as well as monitor and describe the current and emerging tool and the technical platform ecosystem. This will be done in close cooperation with TA1 M2-M4 and other NFDI consortia to ensure their relevance for the users and providers and aligned with existing and emerging national and international initiatives assembled by TA2 M2. Coordinated by GWDG and IPK - and with support of the listed contributors - any useful and relevant tools for biodiversity research might be addressed. The focus, however, is on the development of documentation for the tools and services provided by NFDI4BioDiversity partners. This documentation will target three groups: 1) the broad user community written in clear and simple language (partly in German), 2) technical experts and 3) experienced domain-specific data scientists. Mediated and compiled by co-applicants and participants of NFDI4BioDiversity, it will be made publicly available as published guidelines. The documentation will include 1) data curation/harmonization, data enrichment, quality assurance and control, 2) schema and format validation, 3) improvement of FAIRness of data, 4) reproducibility of experimental data, 5) data publication and data certification, 6) long-term data preservation and archiving of data products and software code, 7) reusability of archived data and data products, 8) code documentation and code repositories for biodiversity tools, and 9) upcoming new demands related to the biodiversity research data management strategy. The last item explicitly comprises tools for the management of research-specific biodiversity workflows to ensure reproducibility of data and the long-term preservation and archiving of upcoming new types of data products in biodiversity research. Screening is done with the methods, tools, mechanisms, rules, conventions and technical services established in GFBio and de.NBI. A knowledge base and a communication platform already exist (on several mediawiki instances of the biowikifarm) and will be extended and adapted. In addition, a new Jupyter instance will be set up in TA1 M4.

Milestones:

Milestones	Description	Month
M3.1.1	NFDI4BioDiversity attribute-value concept for tool documentation established	12
M3.1.2	Training events established together with TA1 M3, TA3 M2 und TA1 M4 and other consortia; a first set of training material available	24
M3.1.3	NFDI4BioDiversity schema for three-fold documentation revised, cooperation with other NFDI consortia and information scientists/ linguistic support started	36
M3.1.4	Knowledge base extension to Jupyter framework adapted and communication platform established together with TA1 M4	48
M3.1.5	Field report and analysis published in peer reviewed journal	60

TA3 M2: Profiling biodiversity tools, technical services and data centers, adopting them as NFDI-RDC-recommended and guide them towards certification

Lead: GWDG; **co-lead:** IPK

Contributors: SNSB, BGBM, UniBremen-MARUM, GFBio e.V., BIBI, HITS, UniGießen, SAB, BGBM-DI, AraGes, GdO, DNAquaNet

Description: Currently, the FAIRness of life science research data and especially their reproducibility is a major topic. This measure will pro-actively promote the profiling, documentation, recommendation and certification of technical services (as established by data repositories) as well as software providing technical assistance in improving data quality at an early stage of the biodiversity research process in the field and bio-lab. Professional profiling of tools, technical services and data pipelines will be done according to the recommended usage in the NFDI4BioDiversity and NFDI-RDC context and according to the envisaged user spectrum (e.g. end-user, data scientist, or server administrator). This is done leveraging GFBio and de.NBI expertise in cooperation with the software developers. To perform the measure, we will establish a „test and certification help desk". This will provide support through experienced bio-software testers, bioinformaticians and data scientists, who are ready to test bio-software, services and repositories. The same helpdesk will guide data centers and software developers towards certification. It will also address the continuous monitoring of all existing NFDI-RDC- and other core services and improving the quality management of NFDI4BioDiversity internal scientific and technical curatorial services. This includes NFDI4BioDiversity internal certification procedures. This helpdesk will closely interact with the user helpdesk established in TA1 M3. Furthermore, this measure will encourage providers of biodiversity tools, technical services and data repositories in biology to use existing national and international technical platforms (described in TA3 M1) for documenting their software, data pipelines, web services, biodiversity ontologies, taxonomies, terminologies (described in TA2 M2) and repositories for data and code. NFDI4BioDiversity will promote open source publication of software. Consequently, this measure will support and guide open-source software developers and providers of technical services to professionalize their implementation and make their documentation publicly available via the recommended platforms. The measure will use JIRA and platforms like GitHub for communication and software documentation, the existing GFBio biowikifarm instances as knowledge base. For further documentation, the measure will use national, European and international platforms like re3data [247], bio.tools [233] and FAIRsharing.org [248], and will closely work together with other NFDI consortia.

Milestones:

Milestones	Description	Month
M3.2.1	Test and certification helpdesk running	12
M3.2.2	Workshop to promote professional profiling, documentation and certification of tools, services and data repositories, together with other NFDI consortia; focus on tools to facilitate the generation of FAIR data	24
M3.2.3	Cooperation with several national, European and international platforms/registries for tools and NFDI-RDC established	36
M3.2.4	Workshop to promote tools and workspaces for managing the whole scientific workflow/ reproducibility of scientific results; together with TA4	48
M3.2.5	Field report and analysis published	60

TA3 M3: Adaptation, Enhancement and Consolidation of NFDI4BioDiversity tool and technical service backbone

Lead: UniJena/iDiv; **co-lead:** SNSB, UniBremen-MARUM

Contributors: BGBM, DSMZ, MfN, SGN, SMNS, SNSB, ZFMK, PANGAEA, Jena Data Center, IPK, GWDG, UFZ, HIFMB, SAB, MPI-BGC, UL, AraGes, BIOfid

Description: Based on the analysis and documentation done in TA3 M1 and TA3 M2 as well as the requirements arising in TA1 and TA4 M1, SIG2 and SIG5 this measure aims to ensure a stable, efficient, future-proof and NFDI-RDC-compliant service and tool backbone that provides FAIR and professional long-term data preservation, publication, and dissemination. This primarily includes work to render the tools, technical services and data pipelines as part of the NFDI-RDC. It also requires their enhancements and consolidation along the steps of the data life-cycle.

Data mobilisation and management: Together with other NFDI consortia, in particular, NFDI4Earth, NFDI4Agri and NFDI4Chem, UniJena, SNSB, and UniBremen-MARUM will develop infrastructures to foster the usage and adaptation of TA3 M2-recommended tools for early mobilisation and management of data. We will extend 1) these tools with emphasis on data reproducibility, efficiency and FAIRness of stored data products, and 2) the data submission, archiving and publication services of NFDI4BioDiversity, with Web APIs for direct NFDI-RDC integration of data to be preserved. This will result in higher data quality and standardised data products and reduce submission, ingest and archiving efforts on both the providers' and the data centers' side. Such data products are easier to integrate in the NFDI-RDC environment (TA4) and last but not least facilitate the OAIS [249] standard-compliant archiving of well-structured re-usable data products and the reproducibility of research results.

Experimental design, provenance and reproducibility: In large joint research projects, e.g. monitoring projects, there is an urgent need to manage biodiversity study design data and scientific workflow data during all stages of the data life-cycle. Tools which address this issue will be amended and adapted for the submission service.

Coordinated by UniJena and SNSB, the data centers will extend the software of the archiving and publication backbone as well as the VAT tool to ensure that data products will be preserved and archived together with the scripts and other software artefacts used to create them. In addition, we will engage in NFDI-wide activities to ensure the preservation of execution environments thus guaranteeing future executability of these scripts, if this is deemed necessary. This measure will also address data provenance in the world of linked open data (LOD) and use and adapt biodiversity identifier registries as e.g. currently implemented by the CETAF and GBIF community. This is done in cooperation with TA2 and TA4. Concerning data provenance in scientific workflows, SNSB and UniJena will amend DWB and BEXIS2 internal features to track, describe and publish coherent working processes.

Schemas and APIs for NFDI-RDC: We will create new types of well-curated highly structured data products. For this goal, we use and develop appropriate schemas for data processing like MOD-CO [123] as well as domain-specific agreed community standards together with TA1 and TA2 as well as other NFDI consortia. All schemas will refer on the schema.org [61] vocabulary for access of structured data on the Internet and follow the W3C design of Linked Open Data (LOD) [250]. Similarly, APIs of data analysis tools will be developed that allow NFDI4BioDiversity compliant submission of data and its integration in the NFDI-RDC environment (TA4 M1).

Dynamic datasets: In addition to the steps described so far the archiving and submission services will need considerable extensions and potentially redesigns to deal with a new class of data. Today's public archives and data repositories run by libraries, publishers and universities focus on static datasets. Especially in biology, we observe a growing number of dynamic datasets with regular updates, e.g. collection and occurrence data or sensor data, and streaming data. We expect these data types to gain even more importance in the future, e.g. in the context of DNA barcoding and citizen science inventories as well as automatic monitoring schemes. Here, new ways for archiving, versioning, and potentially also filtering and aggregation need to be developed based on the first few steps done by biodiversity specific GFBio data centers. This will be a joint activity of the participating data centers coordinated by SNSB.

Landscape consolidation: This measure will also address the consolidation of the landscape of domain-specific data centers and data repositories. Based on the analysis, profiling and certification done in TA3 M1 and M2, UniBremen-MARUM will identify gaps as well as redundancies in the archiving backbone. In cooperation with the "Resource Provider Forum" established in TA4 M5 we will aim to agree on clear responsibilities of the participating data centers and data repositories and clear rules of participation. Ideally, this will result in an archiving backbone network with NFDI-RDC-compliant services (see TA4 M1) linked by identifiers that covers all relevant data, provides replicated storage for each of it, professional archiving infrastructure and avoids unnecessary, costly redundancy in expertise and services.

For this goal, the co-participants and partners will continuously professionalize and adapt their services and make them NFDI4BioDiversity standard and NFDI-RDC compliant.

Operation of services: NFDI4BioDiversity will keep a number of existing tools and platforms technically up-to-date, most of them already in productive use as established in the GFBio project. Their basic operation is contributed in-kind by the NFDI4BioDiversity co-applicants. Additional archiving and publication services, e.g. for new data products and new data types, additional new JIRA installations, additional new code repositories and installations of evolving new analysis tools as well as the adaption of the GFBio BMS monitor service [251] are part of this measure. In case of positive evaluation in TA3 M2, they will be technically expanded by the leading co-applicants and participants. For all of these developments, we will explore whether integration of the tools in the Research Data Commons makes sense. We will ensure that wherever applicable, APIs are constructed in such a way that they can consume data out of the NFDI-RDC and export data into the NFDI-RDC. All APIs and the functionalities underlying them will be described in a way that discovery and semi-automatic composition of pipelines become possible.

Milestones:

Milestones	Description	Month
M3.3.1	First workshop with involvement of certificated GFBio data repositories/ data archives and evolving new NFDI4BioDiversity data repositories done together with TA4 M5 Resource Provider Forum"; landscape consolidated and first additional services operational	12
M3.3.2	First enhancement of data pipelines at data repositories towards schema.org and NFDI-RDC compliance done; first optimized highly structured data products for dynamic data sets	24
M3.3.3	Concept for mobilising lab management, experimental design and assay data as well as scientific workflow data in context with LOD approaches and in accordance with FAIR principles published	36
M3.3.4	The structuring of data products with new data types according evolving new schemas realised in a prototypic version; first optimized highly structured data products for experimental design and provenience data (to facilitate reproducibility of biodiversity research results)	48
M3.3.5	Additional archiving and publication services and Web APIs implemented	60

4.5 Task Area 4: Data integration, exploration, and exploitation - the NFDI-Research Data Commons (NFDI-RDC) (4all & 4future)

Lead: UniMarburg; **co-lead:** IPK, UniBremen-MARUM, UniGießen, HITS

Aims & overview: This task area develops a cloud-based research infrastructure that provides scientists, users and providers with the technical foundation for exchanging data and services in a collaborative research environment while hiding technical aspects as much as possible. This facilitates the integration, use, and sharing of biodiversity data as well as the joint development and reuse of data and services.

The design of the NFDi4BioDiversity Research Data Commons (NFDI-RDC) is aligned with the guidelines of the Open Science Commons of the European Open Science Cloud (EOSC) [131][252] and key concepts of the Australian Research Data Commons [253].

The infrastructure will be implemented together with other NFDI consortia, primed by NFDI4Earth and NFDI4Agri. NFDI-RDC offers on-demand access to scalable compute resources and obeys the function shipping paradigm to avoid time-consuming transfers of large data sets, Figure 4.5.1 outlines the overall architecture of NFDI-RDC. The core of the NFDI-RDC is its mediation layer, addressed in measure TA4 M2. It features common (meta)data models and standards, as well as consistent transformations between these models. The methods for data ingestion either submitted by the user or loaded from an external resource are the main subject of measure TA4 M1. TA4 M3 outlines our storage approach. The raw data is kept in the storage cloud where a metadata index exists for supporting effective searches. Via a transformation process expressed in the mediation layer, the raw data is delivered in a semantic representation most suitable for a target application. The semantic storage layer comprises these representations for the different user communities. TA4 M4 is dedicated to the development of services and applications to foster collaborative research. In addition to data, scientific workflows are also treated as managed resources in NFDI-RDC. Finally, TA4 M5 addresses the management of NFDI-RDC. We introduce agile governance, offer authorization and authentication, and provide monitoring for both the technical infrastructure and the user behaviour. The development and operation of NFDI-RDC will be coordinated and monitored by the NFDI-RDC coordinator at GFBio e.V.
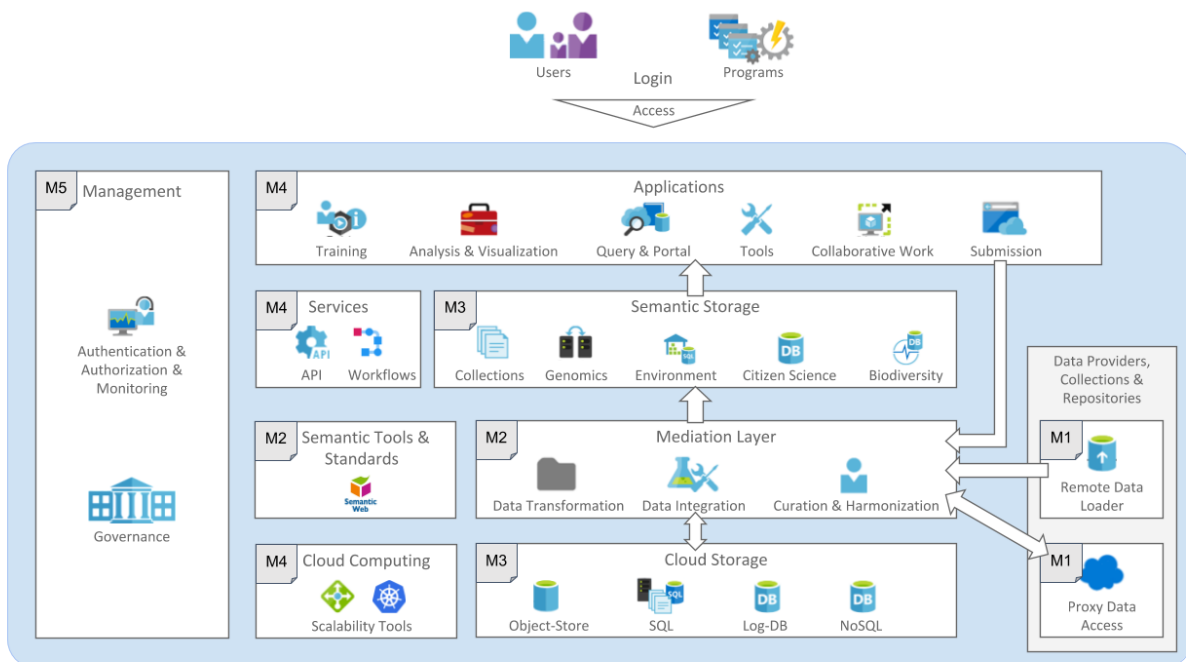

Figure 4.5.1: Overall architecture of the NFDI-RDC. The notation in the upper left corner of the boxes refers to the respective measures (M1-5) within Task Area 4.

TA4 M1: Linking data and service providers with the NFDI-RDC

Lead: UniBremen-MARUM; **co-lead:** GFBio e.V., SNSB

Contributors: InfAI, BGBM, UniGießen, UniMarburg, DSMZ, UniJena/iDiv, BGBM, SUB, IÖR, JKI, BSH, NP Bayerischer Wald

Description: NFDI4Biodiversity's goal is to make the variety of biodiversity data from a multitude of sources available via unified interfaces. Data sources can either be existing data repositories, event streaming data or data directly submitted by users. Data should be accessible by using widely accepted standard formats and searchable based on well-defined and standardized metadata schemas over unified REST interfaces. The data will be made accessible in two ways, either by importing external data into the NFDI-RDC or by providing data proxies. To import data, we will establish a flexible and scalable Extract-Transform-Load (ETL) process.

Import (Extract) of external data: The interfaces of most data sources are very diverse and often require specific implementations to access the data. To reduce the implementation overhead, we will develop guidelines for a REST interface that can be implemented by individual data providers. The data can then be accessed using a generic data access implementation without data source specific components. The data transferred over this interface will be based on the standards developed in TA4 M2 and can be handled using its tools. The design of this interface will evolve from the existing network of data providers and services in GFBio. Based on the developed guidelines new data providers can create compliant implementations of the interfaces to import their data. Since not all data providers will be able to implement this interface, we will supply special tools for accessing non-compliant data sources. The specification for these interfaces will be developed in close cooperation with TA1 M1&2, TA3 M2, TA3 M3, TA4 M2, TA4 M3 and SIG1&2 to achieve broad acceptance.

Transform-Load process: For the transformation step we will use the tooling and standards developed in TA4 M2 to extract metadata from the imported data and transform the data into a suitable standard format if required. The transformed data and metadata will then be loaded into our storage layer using the described REST interface (TA4 M3). The implementation details of this ETL process will be determined in close collaboration with TA4 M2 and TA4 M3.

Data proxies: Not all data can be imported into our system, for example, large next-generation sequencing data sets from ENA [254]. To offer standardized access to these data sources we will develop proxies that deliver links to these databases through standardized REST APIs. Optionally, the requested data will be transparently transformed based on the work in TA4 M2. In addition, these proxies can come with caching mechanisms as known from GFBio, e.g. to keep replicas of metadata in cloud storage.

Milestones:

Milestones	Description	Month
M4.1.1	Guidelines for the REST interface are available and a first reference implementation for a provider exist	12
M4.1.2	Data proxies are available and all data providers from GFBio are supported	24
M4.1.3	Special tools for accessing non-compliant data sources are developed	36
M4.1.4	Entire workflows including ETL processes are implemented and successfully tested with various data providers	48
M4.1.5	Dedicated caching mechanism for data proxies are developed and used in various use cases	60

TA4 M2: Data integration & harmonization

Lead: InfAI; **co-lead:** BGBM

Contributors: UniMarburg, UniBremen-MARUM, UniJena/iDiv, SNSB, SUB, IPK, FBN, HITS, DSMZ, HIFMB, SAB, ZMT, IGB, MPI-Biogeochem, JKI, BSH, AWI, Naturgucker. LAU, BGBM-DI, GdO-Libellen, Arages-Spinnen, NetPhydD-Pflanzen, DDA-Vögel, IFL, NP Hunsrück-Hochwald, LfULG, UniLeipzig

Description: NFDI4BioDiversity aims to build services largely on internationally agreed semantic standards. These standards span both data and corresponding metadata. If such standards are missing or insufficient, especially for interdisciplinary use cases that potentially span multiple NFDIs, additional standards will be developed, in close cooperation with the community, to match their requirements. As detailed in TA4 M3, there is a unique internal representation of the data that is searchable through its associated metadata. Because the data is still not in the required representation of a user community, a mediation layer is introduced to provide the integration, transformation and harmonization for mapping it to the required user representation. These suitable representations for the user communities are offered in the so-called semantic storage layer.

Core models: Developing a generic, common metadata exchange format is a key task of this measure. We will expand the existing GFBio catalogue metadata format to include the extended set of data in NFDI4BioDiversity like livestock and crop data. We will link our efforts with other NFDIs like NFDI4Earth and NFDI4Agri and European initiatives and infrastructures through e.g. ENVRIFair and ELIXIR (see TA2 M3). The definition of the common format will be an extension of Dublin Core and use the schema.org standard. In addition, we will enable the usage of community-accepted standards through the semantic storage layer. Furthermore, we will define schemas for data using similar techniques. A special focus will be the development of exchange formats for numeric environmental or tabular data in general. The defined data and metadata formats will be stored and managed in a component that 1) compiles and annotates concepts, 2) establishes links with semantically related concepts, 3) makes concepts referenceable via stable URIs and 4) makes concept versioning transparent.

Terminology services: The availability of terminologies and ontologies is a requirement for standardization and interoperable analysis of research data. The GFBio terminology service provides a unified API for terminologies needed in the context of national biodiversity research, semantic data annotation and integration (TA3 M3). It offers the ideal basis for the expansion of the service for NFDI. The following specific challenges are addressed: 1) The Terminology Service performance depends in part on the capabilities of decentralized services. This bottleneck is to be overcome by setting up a caching system that centrally stores essential resources for quick execution of queries. 2) Procedures for versioning ontological concepts (incl. provenance annotation) and synchronization with use in distributed software systems have not been sufficiently coordinated so far. Methods, as well as services, required for implementation are to be developed. 3) The potential power of ontologies for the interoperability of research data unfolds through the availability of "mappings" between semantically related concepts used in different community standards. Heuristic methods must be developed to create effective links. 4) The development of terminologies in the research context requires a pragmatically functioning consensus-building process with all actors involved. A model is to be agreed in cooperation with other NFDI consortia (TA2 M1). 5) To facilitate the integration of the terminology services in terms of semantic annotations and data ingestion, tools need to be developed based on specific consumer needs (TA1 M2).

Tools for meta(data) transformation & harmonization: We will reuse existing work (e.g. GFBio Data Transformation Service) as well as relevant terminology resources (e.g. taxonomy and ontology) available through the NFDI-RDC Terminology Services to develop a toolkit for extracting and transforming metadata and data harvested through TA4 M1, and validating the transformed data against the required target schema. The source metadata schemas and data formats used by the data providers will be identified based on an analysis done in TA2 M3. The target schemas will follow the standard models from TA4 M2 and internationally recognized standards in Biodiversity such as DarwinCore, EML and ABCD, and ISO 19115/19139 for geographic metadata. The standard data formats used in the Environmental Sciences such as NetCDF or HDF will also be supported. To meet the wider user needs, the tools will also support the conversion of datasets into easy-to-consume formats (e.g. CSV or JSON). Depending on the data provider requirements, the meta(data) transformation may be carried out either at the data center level using a common pool of meta(data) transformation support tools (e.g. XSLT, OpenRefine) or via a dedicated proxy implementation (see TA4 M1).

Linked Data principles, Biodiversity Knowledge Graph, Semantic Web technologies: In recent years, a growing number of biodiversity data have been published as Linked Open Data (LOD) to support interdisciplinary integration of data.

We will evaluate the potential for interdisciplinary linking of data sets through LOD mechanisms in cooperation with other NFDI consortia and demonstrate its applicability with interdisciplinary pilot applications.

Based on the pilot applications, the required services and APIs will be coordinated and specified. Cooperation with existing international initiatives, such as DiSSCo and iBOL will be important (coordinated by TA2).

Milestones:

Milestones	Description	Month
M4.2.1	Storage solution with caching mechanisms for decentralized resources and common metadata exchange and cataloguing core models published	12
M4.2.2	LOD publication and linking mechanisms and semi-automatic mapping and linking capabilities up and running	24
M4.2.3	Versioning and synchronisation mechanisms implemented	36
M4.2.4	Consensus-building process for terminologies and services implemented	48
M4.2.5	Annotation, adaptable semantic integration and harmonization tools implemented and services running	60

TA4 M3: Storage infrastructure

Lead: UniGießen; **co-lead:** UniMarburg

Contributors: GWDG, BIBI, UniJena/iDiv, SNSB

Description: Recent advances in acquisition and analysis of biodiversity data such as vast image capturing and the application of (meta-)genomics or imaging analysis have resulted in enormous amounts of data, generating a high demand for sustainable and scalable data storage resources. This measure will establish the storage infrastructure required for 1) the operation of data and service providers (TA4 M1), 2) the systematic collection and distribution of metadata required for data integration (TA4 M2), as well as 3) the acquisition of additional data types arising within the context of the application framework (TA4 M4).

Core storage: The core storage layer is designed to offer a scalable and performant, yet flexible and generic storage solution. A wide variety of solutions is available to comply with these requirements: For large raw binary objects, object storage can be used based on the de-facto standard S3 [255]. S3-compatible interfaces allow direct and performant access to the stored objects using standard http calls that can be massively parallelized. For metadata and document-oriented storage NoSQL databases, like MongoDB [256], will be used. For highly structured biodiversity data with fixed and established formats, like ABCD, PostgreSQL [257] with its powerful geo-spatial properties will be applied. Moreover, we offer log-based database systems, e.g. InfluxDB [258], for managing time-series and streaming data sources, which are found in many of our use-cases (TA1 M1), and that are also required for the entire log data items obtained from monitoring the NFDI-RDC infrastructure (TA4 M5).

All these storage components are well-established tools that are under active development and freely available, which minimizes the overhead for maintaining and deploying them.

Access to data: Access to the data layer will be possible via JSON over REST [259]. Design and engineering of the REST API will be conducted in close collaboration with all biodiversity communities. The technical implementation will be based on widely adopted international standards for REST APIs (see also TA4 M1) like openapi 3.0 [260], to make the APIs easily accessible for application developers and data analysts. In addition, the REST API will include versioning of datasets. Depending on the application that uses this API, versioning information can be used either directly to produce provenance information for the end-user or in an indirect manner, e.g. by mapping to external versioning information that is already available. The REST interface will be built on top of Kubernetes which allows easy deployment, scaling and a high degree of portability due to its advanced container management. This will be used to deploy components of the storage platform across multiple sites and ensure thereby that all utilized storage components feature geographically redundant data storage by default.

Semantic storage: Besides applications that directly interact with the cloud storage layer, UniMarburg will implement semantic storage solutions that offer a domain-specific view and advanced features, like dedicated workflows, linked data (TA4 M2) or APIs with specific protocols or schemas, for individual data sets. The data linkage and schema transformation will be based on the tooling developed in TA4 M2. It is currently planned to initially implement semantic storage solutions for 1) biodiversity and collection data, 2) citizen science data (living atlas backend, use case 23), 3) monitoring data (reflecting NRT data + archived data) [261], and 4) omics and imaging data with a focus on biodiversity. The semantic storage solution will be implemented in direct collaboration with the individual communities to offer storage solutions that are optimized for their needs.

Milestones:

Milestones	Description	Month
M4.3.1	Workshop to discuss the initial design of the REST API and finished deployment of basic storage components (MongoDB, PostgreSQL)	6
M4.3.2	Prototype implementation of cloud-based storage infrastructure and first draft requirements for the semantic storage implementations	18
M4.3.3	Working implementations of the semantic storage solutions	30
M4.3.4	Scalable storage solutions available	48
M4.3.5	Final implementation of scalable storage infrastructure on multiple sites	60

TA4 M4: Software application frameworks

Lead: IPK; **co-lead:** UniMarburg, UniGießen

Contributors: BIBI, UniBremen-MARUM, SNSB, SUB, HITS, GWDG, AWI, BSH

Description: This measure provides the frameworks and services to enable scientists to conduct data-driven research on the highest level and in a collaborative manner within NFDI-RDC. In addition to essential FAIR interfaces, it offers powerful scalable tools for processing, analysis, and visualization, as well as generic tools to foster collaborations among researchers. Services are hosted in the cloud storage to facilitate their usage for researchers without technical and administrative issues. Continuous training (TA1 M4, SIG3) will empower users with various skills and scientific background to take full advantage of these frameworks and services in their daily work.

Containers: The whole software backbone is founded on containers and tools like Kubernetes for management and provisioning of containers. This gives the following advantages: 1) containers offer a common lightweight and flexible virtualization that overcome the problem of hard- and software dependencies, 2) Kubernetes facilitates their deployment in a way that users are empowered to deploy their own tools, 3) the complete state of containers can be stored and reloaded again to ensure transparency and reproducibility. The initial software development and deployment will be done at the UniGießen and BIBI site of de.NBI, but a multi-site-setup is always enabled together with other cloud providers like GWDG. Based on the requirements of NFDI4BioDiversity's user communities, preconfigured containers with specific tools are provided. Among others, we support fast and efficient access to Jupyter Notebooks based on the native Kubernetes support of Jupyterhub [262].

Discovery: For data and service discovery, the functionality of the existing GFBio portal is extended to enable the search of services such as workflows that are associated with data sets. This leverages the reuse of services for NFDI users and the international community. In addition, generic software solutions are offered to facilitate the deployment and customization of community portals (use cases 5, 8, 9, 12, 14, 15, 16, 17, 18, 19, 22, 23) with unique domain-specific search features. Semantic searches for data sets and services are enabled based on a full-text search engine. Recommendations derived from the statistics collected in TA4 M5 are given to users to improve the quality of the search results.

Applications: At the application layer of NFDI-RDC, there are powerful tools and workflows for data visualization and analysis in the biodiversity domain that offer expressive visual and analytical processing capabilities with respect to at least three dimensions: time, space, and species. To support high-throughput processing on large datasets, the scale-out of modern cloud computing environments like Apache Flink [263] or Apache Spark [264] is exploited. To scale up, dedicated hardware like GPUs are used for data analysis, especially for machine learning and deep learning approaches, based on Kubeflow [265].

Collaborative work: Collaborative workspace environments will allow joint access to datasets, analysis workflows and results for a defined group of users.

DFG

As the usage of collaborative tools and resources in NFDI-RDC requires specific skills, educational tutorials and training courses are hosted for software developers, data users/providers and data analysts through a collaborative development platform. In addition, we establish a user forum with direct communication channels to experts in the field as technical counterpart for TA1 M4 that focus on the production and organization of training materials.

Milestones:

Milestones	Description	Month
M4.4.1	Design of elastic compute infrastructure finished	6
M4.4.2	Access to collaborative workspace environment established	12
M4.4.3	Training and support platform established and operational	24
M4.4.4	Data and service discovery portal available	36
M4.4.5	Integrated tools for analysis and visualization available	48
M4.4.6	Full access to scalable tools and workflows via containers	60

TA4 M5: Governance & monitoring

Lead: GFBio e.V.; **co-lead:** BIBI, UniBremen-MARUM

Contributors: GWDG, IPK, UniMarburg, UniGießen

Description: This measure will set the provisions for the governance, monitoring, and maintenance of the cloud-based NFDI-RDC, in particular for technical resources and services provided by participants, as well as associated partners. Moreover, unified authentication and authorization of users is an indispensable prerequisite for these actions. All the actions detailed in the following are done in close cooperation with TA5 M1, SIG5 and other NFDIs.

Rules: The first action defines 'Rules for Participation' describing prerequisites, rights, obligations and accountabilities for service and resource providers contributing to NFDI-RDC. These rules will comply with community-specific requirements, legal frameworks including licensing and intellectual property rights (IPR), and the FAIR data principles. A 'Resource Provider Forum' (RPF) consisting of NFDI funded resource providers and third-party providers will be established as part of the cross-cutting topics to manage the rules for participation and to detect new emerging needs like third-party resources. RPF ensures adequate knowledge management, access to knowledge, and protection of intellectual property. This also includes the protection of intellectual property rights in integrated third-party components and the review of the licenses and contracts involved.

AAI: The second action will develop a uniform infrastructure for authentication and authorization to simplify and standardize usage of the diverse resources in a heterogeneous research landscape of the entire NFDI and beyond. To this end, we will employ ELIXIR AAI [266] that was developed in the ELIXIR project and has received great acceptance in both user community and service providers since it was launched in 2016, for example within the de.NBI network coordinated by BIBI.

Since ELIXIR AAI is based on the established OAuth protocol, it is easy to integrate into other NFDI components. It is the foundation for monitoring resource consumption and actions of users within NFDI-RDC.

Control Center: The third action aims at establishing a control center of the technical infrastructure to measure the quality, the resource usage, and other important statistics on services and workflows. For the sake of comparability, data will be managed as time-series in dedicated systems like InfluxDB [258]. Visual dashboards like Grafana [267] allow fast (visual) identification of bottlenecks in the technical resources that are discussed in the RPF to suggest practical actions for improvements.

The last action addresses the monitoring of users regarding their usage of data and services. It will be carried out in full compliance to the General Data Protection Regulation (GDPR) [268] and guided by TA1 M5. Due to the transparency on used resources, it allows to quantify the impact of users on the NFDI and NFDI4BioDiversity. For that, we strive to develop a framework of common and widely accepted research metrics (SIG2) that can be used as both incentives for the users as well as success indicators for funding agencies and universities.

Milestones:

Milestones	Description	Month
M4.5.1	Initiation of resource provider forum finished	12
M4.5.2	Rules for participation established	24
M4.5.3	Authentication and Authorization available	36
M4.5.4	Technical monitoring and performance measurements available	48
M4.5.5	User monitoring and research metrics established	60

4.6 Task Area 5: Coordination, Collaborative governance & sustainability

Lead: GFBio e.V.; **co-lead:** UniBremen-MARUM, DSMZ

In this task area the governance model and coordination of NFDI4BioDiversity will be implemented following the governance model proposed in section 2.5. This includes sustainability aspects and long-term maintenance.

TA5 M1: Coordination and collaborative governance

Lead: UniBremen-MARUM; **co-lead:** GFBio e.V., DSMZ

Contributors: All

Description: Based on the experiences gained with running the GFBio project with 20 partners for seven years, a collaborative coordination, communication, and governance model has been set up as outlined in section 2.5. GFBio e.V. is the key component in this respect since it delivers a dedicated executive anchor to coordinate and manage a large-scale consortium with changing participants in an evolving ecosystem of communities and the emerging NFDI.

GFBio e.V. as an independent legal entity has the further advantage that the hiring procedure is much faster and permanent contracts can be given to avoid the constant loss of expertise. With the tight market especially in the field of data managers, data scientists, IT and software developers this is crucial to be able to have the right experts in place for NFDI. It is also part of the sustainability concept since the success of NFDI will depend on the stability of its human resources. GFBio e.V. will hire 11 FTEs which will be responsible to professionally and collaboratively manage NFDI4BioDiversity internally and as part of NFDI. Additionally, five FTEs will be in charge of IT and software development, for details and duties see below. With its newly-established legal department, DSMZ will contribute their expertise on legal aspects ranging from the cooperation agreement, CBD and Nagoya, to SLAs and contracts/liability.

In general, the applicant, co-applicants and participants of NFDI4BioDiversity would like to state, that the success of NFDI will depend on the possibility to hire experienced staff developing and providing services. The term "Postdoctoral researchers" used in the DFG tables is therefore misleading.

Despite the advantages of making use of GFBio e.V. as outlined above, discussions with UniBremen-MARUM have been started to either directly hire or later transfer the majority of FTEs to UniBremen-MARUM. In case of a positive indication that NFDI4BioDiversity will be funded, the whole "package" of sustainable positions, hiring procedures and rooms will be negotiated. In any case, UniBremen-MARUM has already indicated their strong interest in supporting NFDI4BioDiversity, since engagement in the NFDI is an asset for the city and state of Bremen.

Description of resources:

Executive secretary (Geschäftsführung): The executive secretary will be appointed by the executive board of GFBio e.V. and manage GFBio e.V. on a daily basis. Together with the coordinators and GFBio e.V.'s executive board, they will monitor the progress of NFDI4BioDiversity and suggest any measures to optimise and improve the efficiency in reaching the objectives of NFDI4BioDiversity. He/she will be part of the SC where he/she is involved in the decision making and closely interact with the GFBio e.V. chair and speaker. He/she is responsible for implementing the decisions of the SC and for subsequently reporting any deviations from expectations.

Coordinators: In total six coordinators will be hired: 1) Research Data Commons coordinator. He/She will be responsible to supervise the overall development of the NFDI-RDC platform in TA4. He/she will assist the TAs and MLs of TA4 in their daily business of communication and coordinating the different infrastructures as well as users and providers. 2) Interoperability coordinator. He/she will be responsible for coordinating the exchange and development of metadata standards, terminologies and ontologies within the biodiversity community (national and

international) and between the NFDIs and users/providers. He/she will directly interact and assist TA1, TA2, TA3 and TA4 to enhance interoperability and coordinate the developments. 3) Cross-community coordinator. He/she will coordinate the cross-cutting topics between NFDI4BioDiversity and all other NFDIs as outlined in 2.3. 4) Public relations coordinator. He/she will oversee the public relations activities and user engagement in TA1 and align them with the activities of the different members of NFDI4BioDiversity as well as the NFDI. He/she will chair the SIG1. 5) Service coordinator. He/she will coordinate the service portfolio of NFDI4BioDiversiy and align it with the overall services of NFDI. He/she will take care of implementing monitoring the services in terms of usability and user satisfaction in TA2 and TA3. By leading SIG2 he/she will be involved in the certification of archives as well as quality control and quality assurance. 6) Training and education coordinator. He/she will coordinate the training and education activities in TA1 and TA4 with a focus of initiating the cultural change by awareness building of the value of proper research data management and data literacy in the biodiversity community. He/she will be in close contact with SIG5 and chair SIG3.

IT specialists and Software developers: Two IT specialists and three software developers will be hired by GFBio e.V. acting as a knowledge and implementation task force for the co-applicants, participants and the use cases. Their task is to support the community in setting up their IT and software infrastructure in cooperation with TA3 to reach maximal compatibility with NFDI4BioDiversity. Our experience, by working with the environmental and biodiversity community over the last years, has shown that often relatively small problems constrain the progress in digitalisation and mobilisation of research data. Hiring dedicated IT specialists or software developers for relatively short periods of time is almost impossible for e.g. associations, learned societies or citizen scientists. A flexible task force to help them providing their data in standardized formats will enhance the fairness of their data. After the transition of NFDI4BioDiversity and NFDI from the establishment phase in the operating phase, these IT and software specialists will be well educated and therefore in the prime position to maintain the core services and software components of NFDI4BioDiversity.

Data manager, curator: GFBio e.V. will hire a data manager to assist in data management planning and data curation especially for molecular data. This position will be complementary to the data managers at the collections and repositories by acting also as an information broker and hub for communication.

Administration and Assistance: With GFBio e.V. scaling up from a three-employee association towards a key component of NFDI4BioDiversity with a significantly expanded set of duties and reporting obligations, professional administration structures are a prerequisite.

This will be realised by hiring two FTEs, one for the personnel administration and one for the financial administration, which includes the preparation of the reporting and the communication with the administration of the UniBremen-MARUM. The two persons will be complemented by an assistant who will support the speaker, executive secretary, executive board, steering committee and the board of governors on a daily basis. Tasks range from the organisation of business travels to the wealth of communication and organizational tasks that are typical for large-scale collaborative projects.

Funding: As indicated in the funding request tables NFDI4BioDiversity is ramping up its funding from 2020 to 2025. In 2020 we only request a total of 188,675 Euro allocated to UniBremen-MARUM and GFBio e.V. Together they will act as the "bridgehead" to prepare the start of the NFDI4BioDiversity consortium in January 2021. This includes the preparation of the CA, the joint announcement of the positions with other NFDIs as well as solving all the legal and administrative questions that will arise. Consequently, since the whole consortium starts not before January 2021 the financial request tables extend until December 2025.

Milestones:

Milestones	Description	Month
TA5.1.1	Personnel hired, special interest groups established, Scientific Advisory Board and Board of Governors established	12
TA5.1.2	Collaborative governance and decision workflows established with NFDI directorate	18
TA5.1.3	Evaluation procedures for centrally managed funds established	24
TA5.1.4	Common legal framework for NFDI established	48
TA5.1.5	Transition of NFDI4BioDiversity finished	60

TA5 M2: Sustainability and business model

Lead: GFBio e.V.; **co-lead:** UniBremen-MARUM, DSMZ

Contributors: All

Description: We consider that NFDI4BioDiversity, as well as NFDI, will undergo a steep learning and establishing phase over the next five years. At this point there are too many open questions to establish a sound sustainability concept. The exact duties and obligations of the Directorate are still open and if and when NFDI becomes a legal entity by its own e.g. a not-for-profit NFDI e.V., is not yet decided, to just name two major issues which we expect to be clarified over the next 2-3 years.

As explained in section 2.6 (operating model) NFDI4BioDiversity will start as a classical cooperation project with the UniBremen-MARUM as the applicant GFBio e.V. as the coordinating entity and the co-applicants and participants as the beneficiaries who receive their shares to fulfil their tasks and measures.

This said, we consider that the funds we request for the next five years will be sufficient to not only establish NFDI4BioDiversity, but also leaves enough room for interacting with other consortia as well as the NFDI Directorate to finally build ONE NFDI.

Sustainability: The NFDI4BioDiversity consortium is well aware that an important element in building up data infrastructures is sustainability. For NFDI4BioDiversity this means the maintenance and operation of elaborated technical and organisational structures and services, including adaptations to changing scientific and technical requirements. This has been given thought of from the beginning on. By concept, NFDI4BioDiversity will be implemented as a sustainable infrastructure. NFDI4BioDiversity's funding is complemented by substantial in-kind contributions of personnel and long-term commitments by the co-applicants as detailed in the (Co-) Applicant Contributions in section 6 as well as the letters of commitment of the participants. In-kind contributions cover tasks such as operation and maintenance of hardware and software, licenses, and administrative support. Operation of central services including the helpdesk and data management tasks beyond the respective missions of the participating institutions will be compensated by establishing data management services as an integral and funded part of science projects supported by universities and institutes. This cultural change will be achieved by building up a widespread and cooperative context with science across communities ranging from use cases to front-office/back-office support, education and training as well as reputation for RDM (TA1).

NFDI4BioDiversity's service-oriented architecture, building on modular, re-usable, and exchangeable components further adds to the sustainability of the infrastructure. Used or developed components will be open source and in compliance with international standards. All project related materials and results will be made publicly available. Any source code including documentation will be based in a common repository and open accessible using licenses like Apache 2, LGPL or Creative Commons. Any project developments including high-level comprehensive source code documentation are generally available from the NFDI4BioDiversity website or will be published in appropriate open access media. Furthermore, all NFDI4BioDiversity related data will be citable and can be referenced by persistent identifiers (e.g. Digital Object Identifier - DOI). In summary, even if NFDI4BioDiversity will be decommissioned, only the added value provided by the service infrastructure is lost, but the stored and published data itself are safe.

Despite this, most of the larger infrastructure projects, including federated infrastructures, are facing a problem with sustainability. In general, facilities to be integrated are highly 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 frame and leads to an unbalanced commitment of members.

The speed of IT developments and the ceaseless change in user requirements further complicate the situation. As a consequence, even well-established infrastructures like GBIF partially rely on external grants or - as is the case of many ESFRI projects [269] - necessitate continued centralized funding through EU framework programmes. Even if basic functions of a networked data infrastructure are covered by members, generally available services allowing open data submission and access to integrated data still need additional resources, in particular, personnel. A solution at hand, to compensate for the costs, is to make data management an integral part of research funding, e.g. for science projects or large research equipment, as recommended in a report of the Royal Society [270] and supported by many other organisations including the Belmont Forum [271], the Research Data Alliance [12] and the German Council for Scientific Information Infrastructures (RfII) [272]. However, the practical implementation - as e.g. through funding regulations - is still at an early stage. As indicated above, significant time and effort have to be invested by all NFDI consortia and the NFDI Directorate to join forces for working out pragmatic solutions. NFDI4BioDiversity is prepared for this task by bringing in the experiences from GFBio and de.NBI as well as operating data repositories and information systems for more than 20 years.

Business model: NFDI4BioDiversity will provide a valuable service for providers and users who want to archive, publish and discover data. For the operation of NFDI4BioDiversity and the long-term maintenance of the service components, we will investigate potential sustainability models taking into account the working conditions of the institutions involved. In a first step, NFDI4BioDiversity will define portfolios of services offered by each institution and elaborate a cost model, which allows scientific reuse of data in an open access model. We will also develop cost recovery models for data archiving and publication, as well as for the operation and maintenance of the needed technical platforms. Besides the costs of specific service components, the costs of common services such as portal, training, helpdesk and central user support will be accounted for. Subsequently, a gap analysis will be performed using the results of the cost analysis compared to the available financial basis, which for the establishment phase of NFDI4BioDiversity will be based on the NFDI/DFG funding and in-kind contributions by the project partners. To close any financial gap, we will evaluate additional funding sources like grants and membership fees. The final result will be a common business model laying down the financial rules for the operation of the infrastructure as well as the mode of financial allocations to each contributing component. It should be noted, that the main goal of this measure is to calculate the operational costs of sustaining NFDI4BioDiversity as part of NFDI. We have no intentions to monetize the services or datasets we provide or receive. 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.

A not-for-profit consortium or maybe later on an association is not in the legal position for commercialisation, since besides acceptance problems, liability is almost unsolved. We consider that working out a potential solution to close any financial gap in operating the NFDI will be a collaborative task for the next years to come.

Exit strategy: NFDI4BioDiversity will continuously monitor its performance and acceptance in the community (TA1, SIG2). In case we sense any substantial deviations from expectations either from the users and providers, funders, other NFDIs or the Directorate we will take measures to re-establish a trustworthy and productive working environment and service portfolio. As soon as the first indications of a malfunction of NFDI4BioDiversity appear, the speaker will ask the SC, SAB and BoG for advice. If no solution can be found it will be escalated to the Directorate and the Scientific Senate of NFDI. The speaker will ask the SC to collect all information and make suggestions on how to solve the situation. Together with the SAB, BoG and the Directorate the suggestions will be evaluated and finally the SC will be asked to prepare a basis of a decision on the future of NFDI4BioDiversity for the members meeting of GFBio e.V. If both the SC and GFBio e.V. decide that NFDI4BioDiversiy should be discontinued, the speaker will initiate the decommissioning process by informing all co-applicants and participants as well as the DFG, BMBF, the Directorate and the Senate. Termination of operations is expected until the next regular evaluation of NFDI4BioDiversity, but not longer than five years.

Milestones:

Milestones	Description	Month
TA5.2.1	Sustainability concept and in-kind contributions elaborated	18
TA5.2.2	Service portfolio and cost model established	36
TA5.2.3	Sustainability concept as part of NFDI discussed	48
TA5.2.4	Exit strategy elaborated	60
TA5.2.5	All data and source code publicly available	60

5   List of Abbreviations

Abbreviation	Full name
AAI	Authentication and Authorization Infrastructure
ABCD	Access to Biological Collection Data
AG	Arbeitsgruppe
ALA	Atlas of Living Australia
AMMOD	Automated Multisensor Station for Monitoring of BioDiversity
ANDS	Australian National Data Service
API	Application Programming Interface
ARAMOB	Mobilization of Spider Data: Semantic enrichment and mobilization of data in distributed repositories for taxonomy and ecology
ASA3P	Fully automatic, locally executable and scalable assembly, annotation and analysis pipeline for bacterial genomes
ATLAS	Atlas der Spinnentiere Europas
BiGi	Bielefeld-Gießen Resource Center for Microbial Bioinformatics
BioCASe	Biological Collection Access Service
BLAST	Basic Local Alignment Search Tool
BMBF	Bundesministerium für Bildung und Forschung (Federal Ministry of Education and Research)
BoG	Board of Governors
BOLD	Barcode of Life Data System
BON	Biodiversity Observation Network
BrAPI	Breeding Application Programming Interface
CA	Consortium Agreement
CETAF	Consortium of European Taxonomic Facilities
COST	European Cooperation in Science and Technology
CSV	Comma-Separated Value
DataOne	Data Observation Network of Earth

DFG

Abbreviation	Full name
DCPC	Data Collection and Processing Center
de.NBI	German Network for Bioinformatics Infrastructure
DEEMY	An Information System for Characterization and Determination of Ectomycorrhizae
DFG	Deutsche Forschungsgemeinschaft
DiSSCo	Distributed System of Scientific Collections
DMP	Data Management Plan
DNA	Deoxyribonucleic Acid
DNFS	Deutsche Naturwissenschaftliche Forschungssammlungen
DOI	Digital Object Identifier
DWB	Diversity Workbench
DwC	Darwin Core
DZG	German Centers for Health Research
(e)LTER RI	(Integrated European) Long-Term Ecosystem, Critical Zone & Socio-Ecological Research Infrastructure
e!DAL-PGP	electronic Data Archive Library - Plant Genomics and Phenomics data repository
ECPGR	European Cooperative Programme for Plant Genetic Resources
ECV	Essential Climate Variables
EDGAR	Efficient Database framework for comparative Genome Analyses using BLAST score Ratios
EDIT	European Distributed Institute of Taxonomy
EEZ	Exclusive Economic Zone
ELIXIR	ESFRI infrastructure specialised on bioinformatics.
EML	Ecological Metadata Language
ENVRI	Community for Environmental Research Infrastructures
EnvThes	Environmental Thesaurus
EOSC	European Open Science Cloud
EOSC-Hub	European Open Science Cloud Hub

DFG

Abbreviation	Full name
EOV	Essential Ocean Variables
ESFRI	European Strategy Forum on Research Infrastructures
ESIP	Earth Science Information Partners
ESM	Ecosystem Services and Management
ETL	Extract-Transform-Load
EU BON	European Biodiversity Observation Network
EUDAT	European data infrastructure
EURISCO	European Search Catalogue for Plant Genetic Resources
FAANG	Functional Annotation of Farm Animal Genomes
FAIR	Findable, Accessible, Interoperable, Re-usable
FAIRDOM	FAIR Data Operations Models, transnational project
FAIRDOMHub	FAIRDOMHub is the main installation of the SEEK software, providing data management as a service to systems biologists
FFH	Fauna-Flora-Habitat
FREYA	'Connected Open Identifiers for Discovery, Access and Use of Research Resources', H2020 project on connected, open identifiers
FTE	Full Time Equivalent
GBIF	Global Biodiversity Information Facility
GBIS/I	Genebank Information System of the IPK Gatersleben
GBOL	German Barcode of Life
GDPR	General Data Protection Regulation
GEBA	Genomic Encyclopedia of Bacteria and Archaea
GenDB	Genome annotation system for prokaryotic genomes
GEO BON	Group on Earth Observations; Biodiversity Observation Network
GGBN	Global Genome Biodiversity Network
GHGA	German Genome-Phenome Archive
GOOS	Global Ocean Observing System

DFG

Abbreviation	Full name
GSC	Genomic Standards Consortium
HeFDI	Hessische Forschungsdateninfrastrukturen
HLRN	Norddeutsche Verbund zur Förderung des Hoch- und Höchstleistungsrechnens (High Performance Computing in northern Germany)
I-ADOPT	InteroperAble Descriptions of Observable Property Terminology
iBOL	International Barcode of Life
ICN	International Code of Nomenclature for algae, fungi, and plants
ICNP	International Code of Nomenclature of Prokaryotes
ICOS	Integrated Carbon Observation System
ICSU-WDS	World Data System of the International Science Council
ICZN	International Code on Zoological Nomenclature
iDiv	German Centre for Integrative Biodiversity Research
IDM	Identity Management
INSPIRE	Infrastructure for spatial information in Europe
IPBES	Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services
IPR	Intellectual Property Rights
ISO	International Organization for Standardization
JSON	Javascript Object Notation
KPI	Key Performance Indicators
KR	Knowledge Representation
LIAS	LIAS – A Global Information System for Lichenized and Non-Lichenized Ascomycetes
LOD	Linked Open Data
LPGL	GNU Lesser General Public License
M	Measure
MARLIN	Marine Life Investigator
MARLIN-BSH	Marine Life Investigator-Bundesamt für Seeschifffahrt und Hydrographie

Abbreviation	Full name
MARLIN-S	Marine Life Investigator - Science
MetBan	Automated pipeline for metabarcoding data using taxonomical/phylogenetical classification of organisms
MII	Medical Informatics Initiative
MIxS	Minimal Information about any(x) Sequence
ML	Measure Leader
MOD-CO	An integrative and comprehensive standard for Meta-Omics Data of Collection Objects
MOSES	Modular Observation Solutions for Earth Systems
NABU	Naturschutzbund Deutschland e. V.
NASA	National Aeronautics and Space Administration
NEON	National Ecological Observatory Network
NetCDF	Network Common Data Format
NFDI	Nationale Forschungsdateninfrastruktur
NFDI-RDC	NFDI Research Data Commons
NGO	Non-governmental organization
NOAA	National Oceanic and Atmospheric Administration
NRT	Near Real Time
O2A	Observations to Archives
OAI-PMH	Open Archives Initiative - Protocol for Metadata Harvesting
OBIS	Ocean Biogeographic Information System
OBO	Open Biological and Biomedical Ontology
OGC	Open Geospatial Consortium
OSM	OpenStreetMap
PGR	Plant Genetic Resources
PNU	Prokaryotic Nomenclature Up-to-date
RDA	Research Data Alliance

DFG

Abbreviation	Full name
RDM	Research Data Management
RDMO	Research Data Management Organiser
REST	Representational State Transfer
RfII	German Council for Scientific Information Infrastructures
RPF	Resource Provider Forum
SAB	Scientific Advisory Board
SABIO-RK	System for the Analysis of Biochemical Pathways – Reaction Kinetics
SAP	Systems, Applications and Products
SC	Steering Committee
SDD	Structured Descriptive Data
SIG	Special Interest Group
SIGL	SIG Leaders
sMON	Synthesis project of iDiv (analysing trends in German species data)
SQL	Structured Query Language
SSO	Single Sign On
TA	Task Area
TAL	Task Area Leader
TB	Terabyte
TDWG	Taxonomic Databases Working Group / Biodiversity Information Standards
TERENO	TERrestrial ENvironmental Observatories
TISDAR	RDM toolbox at Thuenen Institute
TYGS	Type Strain Genome Server
USGS	US Geological Survey
VAT	Visualization and Analysis Tool
W3C	World Wide Web Consortium
WCS	Web Coverage Service

DFG

Abbreviation	Full name
WFS	Web Feature Service
WMO	World Meteorological Organization
WMS	Web Map Service

6   Appendix

6.1   Bibliography and list of references

1.   https://www.ipbes.net/global-assessment-report-biodiversity-ecosystem-services

2.   Wilkinson MD, Dumontier M, Aalbersberg IjJ, et al (2016) The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3:160018. https://doi.org/10.1038/sdata.2016.18

3.   https://www.gbif.org

4.   https://cetaf.org/

5.   http://www.eubon.eu/

6.   https://geobon.org

7.   https://www.tdwg.org/

8.   https://lternet.edu/

9.   **https://www.lter-europe.net/elter**

10.   https://www.neonscience.org

11.   http://www.ecpgr.cgiar.org/

12.   https://www.rd-alliance.org

13.   https://www.goosocean.org/

14.   https://www.project-freya.eu

15.   https://data.faang.org/home

16.   https://www.elixir-europe.org/

17.   http://envri.eu/

18.   https://www.eudat.eu/

19.   https://www.dataone.org/

20.   https://www.eosc-hub.eu/

21.   https://www.eosc-portal.eu/

22.   **Glöckner FO, Diepenbroek M, Felden J, et al (2019) Berlin Declaration on NFDI Cross-Cutting Topics. https://doi.org/10.5281/ZENODO.3457213**

23.   **https://www.naturgucker.de**

24.   https://www.re3data.org/search?subjects[]=21 Biology

25.   https://en.wikipedia.org/wiki/Information_silo

26.   https://gfbio.biowikifarm.net/wiki/Data_Publishing/General_part:_GFBio_publication_of_type_1_ data_via_BioCASe_data_pipelines

27.   **Nieschulze J, Astor T (2019) The "long-tail of science" perspective on RDM. Göttingen Research Online / Data. https://doi.org/10.25625/JLHL13**

DFG

28.   https://www.dnfs.de/

29.   https://www.iapt-taxon.org/nomen/main.php

30.   https://www.iczn.org/

31.   http://www.the-icsp.org/bacterial-code

32.   https://www.gbif.org/what-is-gbif

33.   http://www.gbif.de

34.   http://www.denbi.de

35.   http://www.pangaea.de

36.   https://www.pangaea.de/projects/

37.   Kobori H, Dickinson JL, Washitani I, et al (2016) Citizen science: a new approach to advance ecology, education, and conservation. Ecol Res 31:1–19. https://doi.org/10.1007/s11284-015-1314-y

38.   https://www.gbif.org/citizen-science

39.   Chandler M, See L, Copas K, et al (2017) Contribution of citizen science towards international biodiversity monitoring. Biol Conserv 213:280–294. https://doi.org/10.1016/j.biocon.2016.09.004

40.   https://volksbegehren-artenvielfalt.de/

41.   von der Heyde M (2019) Open Research Data: Landscape and cost analysis of data repositories currently used by the Swiss research community, and requirements for the future. https://doi.org/10.5281/ZENODO.2643460

42.   https://www.dfg.de/download/pdf/foerderung/rechtliche_rahmenbedingungen/gute_ wissenschaftliche_praxis/kodex_gwp.pdf

43.   https://www.gfbio.org/gfbio_ev

44.   https://www.hlrn.de/

45.   Mascher M, Schreiber M, Scholz U, et al (2019) Genebank genomics bridges the gap between the conservation of crop diversity and plant breeding. Nat Genet 51:1076–1081. https://doi.org/10.1038/s41588-019-0443-6

46.   https://elixir-europe.org/communities/plant-sciences

47.   https://www.snsb.info/

48.   http://www.lias.net

49.   http://www.deemy.de

50.   https://www.lter-europe.net/elter-esfri

51.   https://www.ufz.de/moses/

52.   https://www.tereno.net/

53.   https://www.digitalearth-hgf.de/

54.   https://www.esm-project.net/

55.   https://www.ufz.de/drp/

56.   https://www.iucn.org/

57.   https://www.ufz.de/teebde/

58.   https://www.ipbes.net

59.   https://www.idiv.de/en/smon.html

60.   www.uni-marburg.de/hefdi

61.   https://schema.org/

62.   https://terminologies.gfbio.org

63.   https://www.icsu-wds.org/

64.   https://dini.de/

65.   https://www.langzeitarchivierung.de/Subsites/nestor/EN/

66.   https://www.coretrustseal.org/

67.   https://www.fairsfair.eu/

68.   https://www.uni-bremen.de/en/research-alliance/

69.   https://gbis.ipk-gatersleben.de

70.   https://www.ilter.network/

71.   http://alter-net.info/

72.   https://www.gbif.org/news/6ewyUhBpRYammYWI2CgsM4/biodiversity-infrastructures-to-crosslink-metagenomics-and-species-occurrence-data

73.   https://www.rd-alliance.org/rda-disciplines

74.   https://www.esipfed.org

75.   https://www.nasa.gov/

76.   https://www.noaa.gov/

77.   https://www.usgs.gov/

78.   Diepenbroek M, Grobe H, Reinke M, et al (2002) PANGAEA—an information system for environmental sciences. Comput Geosci 28:1201–1210. https://doi.org/10.1016/S0098-3004(02)00039-0

79.   Diepenbroek M, Schindler U, Huber R, et al (2017) Terminology supported archiving and publication of environmental science data in PANGAEA. J Biotechnol 261:177–186. https://doi.org/10.1016/j.jbiotec.2017.07.016

80.   http://www.emplantbase.org/home.html

81.   http://biodiversity.eubon.eu/eu-bon-taxonomic-backbone

82.   http://www.biocase.org/

83.   https://www.researchgate.net/profile/Walter_Berendsohn/publication/263083477_The_BioCASE_Project_-_a_Biological_Collections_Access_Service_for_Europe/links/0deec539b72946ed67000000.pdf

84.   https://cybertaxonomy.eu/

85.   https://annosys.bgbm.fu-berlin.de/

86.   http://www.ggbn.org/ggbn_portal/

87.   https://www.denbi.de/literature/de-nbi-special-issue-in-journal-of-biotechnology

88.   https://elixir-europe.org/about-us/who-we-are/nodes/germany

89.   https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud

90.   https://www.eosc-portal.eu/eosc-life

91.   https://bacdive.dsmz.de/

92.   https://bacdive.dsmz.de/api/pnu/

93.   Söhngen C, Bunk B, Podstawka A, et al (2013) BacDive—the Bacterial Diversity Metadatabase. Nucleic Acids Res 42:D592–D599. https://doi.org/10.1093/nar/gkt1058

94.   Söhngen C, Podstawka A, Bunk B, et al (2015) BacDive – The Bacterial Diversity Metadatabase in 2016. Nucleic Acids Res 44:D581–D585. https://doi.org/10.1093/nar/gkv983

95.   Reimer LC, Söhngen C, Vetcininova A, Overmann J (2017) Mobilization and integration of bacterial phenotypic data—Enabling next generation biodiversity analysis through the Bac Dive metadatabase. J Biotechnol 261:187–193. https://doi.org/10.1016/j.jbiotec.2017.05.004

96.   **Reimer LC, Vetcininova A, Carbasse JS, et al (2018) BacDive in 2019: bacterial phenotypic data for High-throughput biodiversity analysis. Nucleic Acids Res 47:D631–D636. https://doi.org/10.1093/nar/gky879**

97.   **https://farmpheno.fbn-dummerstorf.de/FarmPhenoDB.jsf?conversationContext=1**

98.   Telaar A, Liland KH, Repsilber D, Nürnberg G (2013) An Extension of PPLS-DA for Classification and Comparison to Ordinary PLS-DA. PLoS One 8:e55267. https://doi.org/10.1371/journal.pone.0055267

99.   **http://www.gfbio.org**

100.   **Diepenbroek, M., Glöckner, F. O., Grobe, P., et al (2014). Towards an integrated biodiversity and ecological research data management and archiving platform: the German federation for the curation of biological data (GFBio). In: Plödereder, E., Grunske, L., Schneider, E. & Ull, D. (Hrsg.), Informatik 2014. Bonn: Gesellschaft für Informatik e.V.. (S. 1711-1721). https://dl.gi.de/handle/20.500.12116/2782**

101.   **http://www.gwdg.de**

102.   https://www.pidconsortium.eu

103.   **https://fair-dom.org**

104.   **https://fairdomhub.org/**

105.   **Wolstencroft K, Krebs O, Snoep JL, et al (2016) FAIRDOMHub: a repository and collaboration environment for sharing systems biology research. Nucleic Acids Res 45:D404–D407. https://doi.org/10.1093/nar/gkw1032**

106.   **https://rightfield.org.uk**

107.   **Wolstencroft K, Owen S, Horridge M, et al (2011) RightField: embedding ontology annotation in spreadsheets. Bioinformatics 27:2021–2022. https://doi.org/10.1093/bioinformatics/btr312**

108.   **http://sabiork.h-its.org**

109.   **https://edal-pgp.ipk-gatersleben.de/**

110.   **Arend D, Junker A, Scholz U, et al (2016) PGP repository: a plant phenomics and genomics data publication infrastructure. Database 2016:baw033. https://doi.org/10.1093/database/baw033**

111.   Oppermann M, Weise S, Dittmann C, Knüpffer H (2015) GBIS: the information system of the German Genebank. Database 2015:. https://doi.org/10.1093/database/bav021

112.   Ćwiek-Kupczyńska H, Altmann T, Arend D, et al (2016) Measures for interoperability of phenotypic data: minimum information requirements and formatting. Plant Methods 12:. https://doi.org/10.1186/s13007-016-0144-4

113.   Selby P, Abbeloos R, Backlund JE, et al (2019) BrAPI—an application programming interface for plant breeding applications. Bioinformatics. https://doi.org/10.1093/bioinformatics/btz190

114.   Weise S, Oppermann M, Maggioni L, et al (2016) EURISCO: The European search catalogue for plant genetic resources. Nucleic Acids Res 45:D1003–D1008. https://doi.org/10.1093/nar/gkw755

115.   http://www.bayernflora.de

116.   **http://www.diversitymobile.net**

117.   **http://www.mod-co.net**

118.   **https://biowikifarm.net**

119.   http://www.snsb.info/Services.html

120.   **http://services.snsb.info/DTNtaxonlists/rest/v0.1/static/api-doc.html**

121.   **https://diversityworkbench.net/Portal/Software**

122.   **Weiss M, Weibulat T, Seifert S, et al (2018) A flexible Diversity Workbench tool to publish biodiversity data from SQL database networks through platforms like GFBio. Friedrich-**

Schiller-Universität Jena

123. **Rambold G, Yilmaz P, Harjes J, et al (2019) Meta-omics data and collection objects (MOD-CO): a conceptual schema and data model for processing sample data in meta-omics research. Database 2019:. https://doi.org/10.1093/database/baz002**

124. **Triebel D, Reichert W, Bosert S, et al (2018) A generic workflow for effective sampling of environmental vouchers with UUID assignment and image processing. Database 2018:. https://doi.org/10.1093/database/bax096**

125. **Rambold G, Bensch K, Kirk PM, et al (2017) Citation of a taxon name identifier issued by the ICN-recognized registration repositories instead of taxon name author citation. Taxon 66:1200–1203. https://doi.org/10.12705/665.12**

126. **Rambold G, Zedda L, Coyle JR, et al (2016) Geographic heat maps of lichen traits derived by combining LIAS light description and GBIF occurrence data, provided on a new platform. Biodivers. Conserv. 25:2743–2751**

127. **https://www.ufz.de/index.php?de=37234**

128. **https://www.ufz.de/lter-d/**

129. **https://www.uni-giessen.de/fbz/fb08/Inst/bioinformatik/software/**

130. **Meyer F (2003) GenDB--an open source genome annotation system for prokaryote genomes. Nucleic Acids Res 31:2187–2195. https://doi.org/10.1093/nar/gkg312**

131. https://ec.europa.eu/info/publications/european-open-science-cloud-eosc-strategic-implementation-plan_en

132. **https://fusion.cs.uni-jena.de/bpp/**

133. **https://idata.idiv.de**

134. **https://www.bexis.uni-jena.de**

135. **http://www.the-jena-experiment.de/Data.html**

136. **http://www.aquadiva.uni-jena.de/**

137. **https://www.gfbio.org/de/visualize**

138. **Authmann C, Beilschmidt C, Drönner J, et al (2015) VAT: A System for Visualizing, Analyzing and Transforming Spatial Data in Science. Datenbank-Spektrum 15:175–184. https://doi.org/10.1007/s13222-015-0197-y**

139. https://www.uni-marburg.de/de/forschung/kontakt/forschungsdatenmanagement

140. **https://arages.de**

141. **https://atlas.arages.de/**

142. **http://www.aramob.de/web/**

143. **https://www.awi.de/en/about-us/service/computing-centre/data-flow-framework.html**

144. **Koppe R, Gerchow P, Macario A, et al (2015) O2A: A generic framework for enabling the flow of sensor observations to archives and publications. In: OCEANS 2015 - Genova. IEEE**

145. **Hoffmann N, Berendsohn W, Güntsch A, et al (2011) Biodiversity information platforms: From standards to interoperability. Zookeys 150:71–87. https://doi.org/10.3897/zookeys.150.2166**

146. **Proft S, Grau J, Caswara C, et al (2017) Metban: Automated Pipeline For Metabarcoding Data Using Taxonomical/Phylogenetical Classification Of Organisms**

147. **http://www.algaterra.net/**

148. https://www.bgbm.org/sites/default/files/documents/cediatom120Kusber%2BJahn.pdf

149. **https://wiki.ggbn.org/ggbn/Specimen_Tool**

150. **Gemeinholzer B, Dröge G, Zetzsche H, et al (2011) The DNA Bank Network: The Start from a German Initiative. Biopreserv Biobank 9:51–55. https://doi.org/10.1089/bio.2010.0029**

151. **Droege G, Barker K, Seberg O, et al (2016) The Global Genome Biodiversity Network (GGBN) Data Standard specification. Database 2016:baw125. https://doi.org/10.1093/database/baw125**

152. https://www.openstarts.units.it/bitstream/10077/3737/1/Berendsohn%2C bioidentify.pdf

153. **https://www.biofid.de/en/search/**

154. **Koch M, Driller C, Schmidt M, et al (2018) Current progress in the development of taxonomic and anatomical ontologies within the scope of BIOfid. Biodivers. Inf. Sci. Stand. 2:e25585**

155. **Driller C, Koch M, Schmidt M, et al (2018) Workflow and Current Achievements of BIOfid, an Information Service Mobilizing Biodiversity Data from Literature Sources. Biodivers Inf Sci Stand 2:e25876. https://doi.org/10.3897/biss.2.25876**

156. **https://www.bsh.de/DE/DATEN/daten_node.html**

157. **https://epic.awi.de/id/eprint/48140/**

158. http://www.osnabrueck.bund.net/

159. **http://www.ornitho.de/**

160. **http://www.dnaqua.net/**

161. Leese F, Bouchez A, Abarenkov K, et al (2018) Why We Need Sustainable Networks Bridging Countries, Disciplines, Cultures and Generations for Aquatic Biomonitoring 2.0: A Perspective Derived From the DNAqua-Net COST Action. Next Gener. Biomonitoring Part 1 63–99

162. **http://www.libellula.org/**

163. **http://ichthyologie.de/index.php/de/**

164. https://www.gfoe.org/

165. http://www.multibasecs.de

166. **http://www.nationalatlas.de/**

167. **https://www.ifl-leipzig.de/de/forschung/geovisualisierungen.html**

168. https://begendiv.de/

169. https://allianceforfreshwaterlife.org/

170. Darwall W, Bremerich V, De Wever A, et al (2018) The Alliance for Freshwater Life : A global call to unite efforts for freshwater biodiversity science and conservation. Aquat Conserv Mar Freshw Ecosyst 28:1015–1022. https://doi.org/10.1002/aqc.2958

171. **https://wiki.dbpedia.org/**

172. **Bizer C, Lehmann J, Kobilarov G, et al (2009) DBpedia - A crystallization point for the Web of Data. J Web Semant 7:154–165. https://doi.org/10.1016/j.websem.2009.07.002**

173. **https://monitor.ioer.de**

174. **Walz U, Stein C (2018) Indicator for a monitoring of Germany's landscape attractiveness. Ecol Indic 94:64–73. https://doi.org/10.1016/j.ecolind.2017.06.052**

175. **http://www.openagrar.de**

176. **https://www.tierstimmenarchiv.de**

177. https://www.europeana.eu

178. **https://forschungsfallnachtigall.de/**

179. **http://naturblick.naturkundemuseum.berlin/**

180. **Petersen M, Glöckler F, Hoffmann J (2019) Harmonizing plot data with collection data. Res Ideas Outcomes 5:. https://doi.org/10.3897/rio.5.e33509**

181. **https://github.com/esa-esdl/ESDL.jl**

DFG

182. **https://cran.r-project.org/web/packages/dimRed/index.html**

183. **https://www.earthsystemdatalab.net/**

184. **https://deutschlandflora.de/dflor/**

185. **http://recorder-d.de/**

186. https://www.nationalpark-bayerischer-wald.bayern.de/

187. **https://fs.nlphh.de/**

188. **http://www.ingrid-oss.eu**

189. **https://www.gda.bayern.de/service/findmitteldatenbank/**

190. Díaz S, Demissew S, Carabias J, et al (2015) The IPBES Conceptual Framework — connecting nature and people. Curr Opin Environ Sustain 14:1–16. https://doi.org/10.1016/j.cosust.2014.11.002

191. https://www.dissco.eu/

192. **http://www.insekten-sachsen.de**

193. **https://naturportal-suedwest.de/**

194. https://open-access.net

195. Hätscher P (2007) Open Access an deutschen Hochschulen. Institutional Repositories und die Informationsplattform open-access.net. Z Bibliothekswes Bibliogr 55:216–223. https://doi.org/10.3196/18642950085445125

196. http://www.uni-goettingen.de/en/1.html

197. https://www.idiv.de/en/index.html

198. **https://www.morphdbase.de/**

199. https://www.bolgermany.de/

200. Geiger MF, Astrin JJ, Borsch T, et al (2016) 22 How to tackle the molecular species inventory for an industrialized nation—lessons from the first phase of the German Barcode of Life initiative GBOL (2012–2015). Genome 59:661–670. https://doi.org/10.1139/gen-2015-0185

201. https://www.leibniz-zmt.de/en/

202. https://www.ands.org.au/

203. https://www.dfg.de/download/pdf/dfg_im_profil/gremien/senat/biodiversitaet/ag_ biodiversitatsdaten_datensicherung.pdf

204. https://www.cancer.gov/research/nci-role/bioinformatics/cancer-research-data-ecosystem-infographic

205. https://ardc.edu.au/

206. https://www.tdwg.org/standards/dwc/

207. https://www.tdwg.org/standards/abcd/

208. **https://gfbio.biowikifarm.net/wiki/Data_exchange_standards,_protocols_and_formats_ relevant_ for_the_collection_data_domain_within_the_GFBio_network**

209. http://www.dcc.ac.uk/resources/metadata-standards/eml-ecological-metadata-language&gt

210. **Yilmaz P, Kottmann R, Field D, et al (2011) Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. Nat Biotechnol 29:415–420. https://doi.org/10.1038/nbt.1823**

211. https://www.tdwg.org/standards/sdd/

212. https://wiki.ggbn.org/ggbn/GGBN_Data_Standard

213. https://www.openarchives.org/pmh/

214. http://biocase.org/products/protocols/index.shtml

DFG

215.   https://www.w3.org/community/openannotation/

216.   https://cetafidentifiers.biowikifarm.net/wiki/Main_Page

217.   https://diversityworkbench.net/

218.   https://abcd.tdwg.org/

219.   https://dwc.tdwg.org/

220.   https://www.iso.org/standard/32554.html

221.   https://www.gbif.org/standards

222.   https://inspire.ec.europa.eu/

223.   https://www.iso.org/standard/32557.html

224.   https://bioschemas.org/

225.   http://www.obofoundry.org

226.   https://www.go-fair.org/implementation-networks/overview/biodifairse/

227.   https://www.rd-alliance.org/groups/interoperable-descriptions-observable-property-terminology-wg-i-adopt-wg

228.   https://www.rd-alliance.org/groups/rdawds-publishing-data-ig.html

229.   https://www.rd-alliance.org/groups/assessment-data-fitness-use

230.   https://www.rd-alliance.org/groups/fair-data-maturity-model-wg

231.   http://www.copdess.org/enabling-fair-data-project/

232.   Hodson S, Jones S, Collins S, et al (2018) FAIR Data Action Plan: Interim recommendations and actions from the European Commission Expert Group on FAIR data

233.   https://bio.tools/

234.   https://www.ala.org.au/

235.   https://www.gfbio.org/tools

236.   https://www.cbd.int/

237.   https://www.eoscpilot.eu/sites/default/files/fair4s_eoscpilot_skills_framework.pdf

238.   https://dalspace.library.dal.ca/bitstream/handle/10222/64578/Strategies and Best Practices for Data Literacy Education.pdf

239.   http://www.uni-goettingen.de/en/56640.html

240.   https://www.vbio.de/

241.   https://www.gfbs-home.de/

242.   https://www.dzg-ev.de/en/

243.   http://www.biodiversity.de/

244.   https://www.go-fair.org/go-fair-initiative/go-change/

245.   http://www.catalogueoflife.org/

246.   https://www.gfbio.org/data-centers

247.   https://www.re3data.org/

248.   https://fairsharing.org/

249.   http://www.oais.info/

250.   https://www.w3.org/DesignIssues/LinkedData.html

251.   https://bms.gfbio.org/

252.   https://ec.europa.eu/research/openscience/pdf/eosc_declaration.pdf

253.   Barker M, Wilkinson R, Treloar A (2019) The Australian Research Data Commons. Data Sci J 18:. https://doi.org/10.5334/dsj-2019-044

254.   https://www.ebi.ac.uk/ena

255.   https://aws.amazon.com/de/s3/

256.   https://www.mongodb.com/

257.   https://www.postgresql.org/

258.   https://www.influxdata.com/

259.   https://en.wikipedia.org/wiki/Representational_state_transfer

260.   https://www.openapis.org/

261.   https://www.fona.de/de/aktuelles/nachrichten/2019/kabinett-verabschiedet-aktionsprogramm-insektenschutz.php

262.   https://jupyter.org/hub

263.   https://flink.apache.org/

264.   https://spark.apache.org/

265.   https://www.kubeflow.org/

266.   https://elixir-europe.org/services/compute/aa

267.   https://grafana.com/

268.   https://eur-lex.europa.eu/eli/reg/2016/679/oj

269.   https://www.esfri.eu/

270.   https://royalsociety.org/topics-policy/projects/data-governance/

271.   http://www.belmontforum.org/

272.   http://www.rfii.de/

DFG