

Connecting National and International Data Infrastructures in Biodiversity Research

The Case of NFDI4Biodiversity, a German Consortium for Biodiversity, Ecology and Environmental Data

Barbara Ebert¹[\[https://orcid.org/0000-0003-3328-6693\]](https://orcid.org/0000-0003-3328-6693), Judith Sophie Engel² [\[0000-0001-8665-6382\]](https://orcid.org/0000-0001-8665-6382), Ivaylo Kostadinov¹[\[https://orcid.org/0000-0003-4476-6764\]](https://orcid.org/0000-0003-4476-6764), Anton Güntsch³[\[https://orcid.org/0000-0002-4325-4030\]](https://orcid.org/0000-0002-4325-4030) and Frank Oliver Glöckner^{2,4} [\[https://orcid.org/0000-0001-8528-9023\]](https://orcid.org/0000-0001-8528-9023)

¹ GFBio - German Federation for Biological Data, Germany

² MARUM - Center for Marine Environmental Sciences, University of Bremen, Germany

³ Center for Biodiversity Informatics and Collection Data Integration, Botanic Garden and Botanical Museum Berlin, Germany

⁴ AWI - Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Germany

Abstract. NFDI4Biodiversity is a consortium within the German National Research Data Infrastructure (NFDI) dedicated to data and services for biodiversity research and ecology. In the domain, well-developed international networks exist, with quite mature tools and data standards. These are used and disseminated in the NFDI4Biodiversity project in order to mobilise and publish data collected by national stakeholders according to the FAIR Guiding Principles for scientific data management and stewardship. The consortium partners provide methods and tools for archiving, publishing, searching and analysing data that are suitable for everyday use and have been tried and tested in practice. The consortium also functions as a forum for technical and legal matters of data and data-related workflows. In this way, the consortium is providing added value to the community regarding access to modern technologies and a comprehensive stock of biodiversity and environmental data.

Keywords: Biodiversity, Ecology, Community-based services, Research Data, NFDI, Germany, NFDI4Biodiversity

1. Setting the Scene

NFDI4Biodiversity is a consortium within the German National Research Data Infrastructure (NFDI) dedicated to data and services for biodiversity research and ecology [1]. The 50 partner organisations represent typical stakeholders in the domain: Research groups with a variety of methods and data types from taxonomic and ecosystem research, natural history collections with digitisation projects for data and objects, and the nature conservation domain, where Citizen scientists and highly specialised expert organisations contribute large amounts of structured observation data. IT service centres and informatics research groups bring important technical competencies to the group.

Data on the biodiversity of animals, plants and microorganisms typically include species observations recorded in the form of tables, photos, videos, specimens and audio files. Specimens are also frequently collected and preserved for further analysis and description.

Increasingly important is genetic information on the observed species, i.e. sequence data and data from other -omics methodology (metabolomics, glycomics, and transcriptomics), which provide information on biological functions. Contextual data on environmental conditions in the species' habitat, on land use or colonisation are also relevant. Data streams from continuous monitoring of environmental parameters by sensors/satellites bring on new infrastructure requirements.

Well-developed international networks exist, with quite mature tools and data standards (see Table 1). These are used and disseminated in the NFDI4Biodiversity project in order to mobilise and publish data collected by national stakeholders according to the FAIR Guiding Principles for scientific data management and stewardship. NFDI4Biodiversity also builds on services provided by the German Federation for Biological Data (GFBio e.V.), the German Network on Bioinformatics Infrastructure - Deutsches Netzwerk für Bioinformatik Infrastruktur (de.NBI), and the eight German GBIF nodes.

Table 1. International data initiatives with standards and tools of practical relevance to the current work (in alphabetical order)

Name	Scope	Tools and standards relevant in current work
BioCASE - Biological Collection Access Service	Network of primary biodiversity repositories	BioCASE protocol BioCASE monitor software BioCASE provider software
CETAF - Consortium of European Taxonomic Facilities	Object-related and taxonomic research, biological and geological collections	Best practices - CETAF stable identifiers
ELIXIR - European Research Infrastructure for the Life Sciences	Molecular and -omics research	European Nucleotide Archive ENA RDM-Kit - Research Data Management toolkit Bioschemas Policy and specification
GBIF - Global Biodiversity Information Facility	Data, common standards, open-source tools	Scientific Data Collection Hosted portals Integrated Publishing Toolkit
International Barcode of Life	Data and DNA-based tools	Barcode of Life Data System
RDA - Research Data Alliance	Recommendations and standards	I-ADOPT Framework
TDWG - Biodiversity Information standards	Standards and guidelines for the recording and exchange of data about organisms	Access to Biological Collection Data (ABCD) Schema

GFBio services are structured along the data life cycle, from the preparation of a data management plan to data archiving and publishing [2]. A data submission service is provided, where researchers can submit heterogeneous data files from their projects for curation and archiving in matching data archives, including the European Nucleotide Archive ENA. Any submission or consultation generates a ticket that is handled by a professional helpdesk team with representatives from the associated data centres.

2. The NFDI4Biodiversity approach

Mobilising data is facilitated if partners get added value for their own day-to-day business, for example the digitisation of work processes. In a use case project with the Bavarian Forest National Park, BEXIS (Information System of the Biodiversity Exploratories) was established as a hosted database solution for scientific data management in the national park's projects, facilitating data publications in the future. In the use case for a Living Atlas of Germany, partners are provided with BioCASE installations to facilitate data publications from local systems to the GBIF and GFBio data portals. In 2022, flexible funds from the NFDI4Biodiversity consortium were used to create the development of an open web service interface to the new checklist infrastructure of the National Red List Centre.

In order to increase our outreach to scientists, we introduced a "Front Office/Back Office model" in the Helpdesk support: Initial consulting by local data stewards at research performing organisations is linked to the subject-specific consulting services of the NFDI4Biodiversity network (see Fig. 1). The model was prototypically implemented with the Research Initiative for Biodiversity Conservation (FEEdA) [3]. We see the model as an effective building block for the professional management of data from large research collaborations as well as the "long tail of science" at universities and institutes.

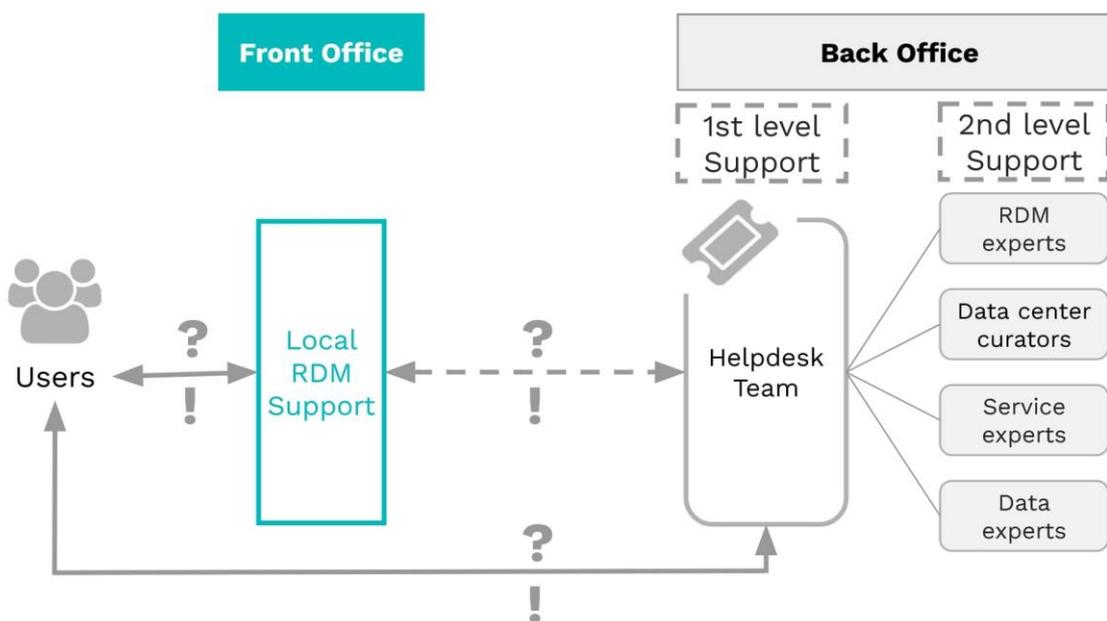


Figure 1. Front Office/Back Office support provided by the NFDI4Biodiversity Helpdesk (adapted from [3])

On a technical level, the vision of NFDI4Biodiversity is a federated, distributed IT infrastructure ("Commons") that provides users with access to tried-and-tested applications for data analysis and publication, in combination with tools for storage and semantic integration. Scientific IT centres in the network provide cloud infrastructure and storage and act as bridges into the network of high-performance computing centres.

3. Conclusions

By combining international tools and standards with a national service infrastructure, the NFDI4Biodiversity consortium is providing access to modern technologies and a comprehensive stock of biodiversity and environmental data. The consortium is also a forum for technical and legal matters of data and data-related workflows. With regard to future

development, we see two important lines of development: a) sustainable funding for the public core data resources in the life sciences; b) integration of data management structures in the national and international biodiversity monitoring programmes with the scientific research infrastructure landscape. In this way, we can mobilise and provide the extensive data that stakeholders from science, politics, economy and society need for better contributions to the conservation of global biodiversity.

Data availability statement

The submission is not based on data.

Underlying and related material

none

Competing interests

The authors declare that they have no competing interests.

Funding

NFDI4Biodiversity is funded by the German Research Foundation (DFG) within the framework of an agreement between the Federal Government and the Länder on the establishment and funding of the National Research Data Infrastructure (NFDI) of 26 November 2018 (grant number [442032008](#)).

Acknowledgement

We thank the dedicated staff and partners who help to shape and support the structures and results presented here. With regard to this article special thanks go to Merle Schwarten for editing the manuscript.

References

1. F. O. Glöckner et al. "NFDI4BioDiversity - A Consortium for the National Research Data Infrastructure (NFDI)", Zenodo, Jul 2020, <https://doi.org/10.5281/zenodo.3943645>
2. German Federation for Biological Data: GFBio Services, <https://www.gfbio.org/services/> (accessed on 2023-04-25)
3. FEdA is a 10-year funding program with 25-30 research projects and a central coordination office (www.feda.bio). For information on the NFDI collaboration see: J. Linares Gómez, B. Ebert, J. Eberhardt, K. Frohne, K. Sauerland, K. Mozygamba, B. Miller. "Collaborative Research Data Management Support for FEdA projects", Zenodo, Feb 2023, <https://doi.org/10.5281/zenodo.7624583>