NFDI4Biodiversity - National and international involvement

Second strategic report

Mai 2023

Authors

Anton Güntsch[1], David Fichtmüller[1], Barbara Ebert[2], Kristin Sauerland[2], Dagmar Triebel[3], Martin Zurowietz[4], Mark Frenzel[5]

[1]Freie Universität Berlin, Botanic Garden and Botanical Museum Berlin

[2]German Federation for Biological Data (GFBio e.V.)

[3]Staatlichen Naturwissenschaftlichen Sammlungen Bayerns

[4]Universität Bielefeld

[5]Helmholtz Centre for Environmental Research UFZ

Content

Executive Summary

Based on the "1st strategic report on national and international involvement", we have started to plan and implement measures for increased interactions with the prioritized actors. These are, depending on the target infrastructure, very manifold and range from concrete implementation measures to the organisation of events under the umbrella of the respective target infrastructure to the formulation of Memoranda of Understanding (MoU). Out of 10 prioritized target infrastructures, we have already strengthened our involvement with 8 through various activities described in the report.

So far, Research Data Alliance (RDA) and the German Centre for Integrative Biodiversity Research (iDiv) have not been specifically addressed even though there are strong links through existing individual activities. These will be approached and structured in the next phase of the NFDI4Biodiversity project.

1 Introduction

In its first "Strategic report on national and international involvement", the NFDI4Biodiversity Consortium defines 10 actors in the national and international biodiversity data infrastructure landscape that are of outstanding importance for a German national infrastructure [1]. These were identified in a participatory process, initially by consulting the entire consortium, and prioritized in the Special Interest Group "Strategy".

The resulting priority infrastructures and projects were 1) GBIF, 2) BfN, 3) RDA, 4) EOSC, 5) Bioschemas.org, 6) ALA, 7) de.NBI, 8) FEdA, 9) iDiv, and 10) TDWG. The preliminary identification of target infrastructures was accompanied by the creation of dynamic documentation on the NFDI4Biodiversity Confluence platform, where priorities can be calibrated at any time and existing activities can be documented. The documentation platform has been continuously expanded and provides project partners with a valuable tool for describing essential external relations. Besides the ten infrastructures and projects with the highest priority, over 60 other initiatives of medium or low priority are documented there as well.

Since the preparation of the first strategic report, there have been two changes to the list of infrastructures and projects with particularly high priority: On the one hand, a technical assessment of the platform "Atlas of living Australia" (ALA) revealed that the originally planned use in the context of the NFDI4Biodiversity Use Case "LAND" (Lebendiger Atlas Natur Deutschland) will not be realised for the time being. ALA is therefore no longer one of the high-priority infrastructures at the moment. On the other hand, it was decided that the "Gesellschaft für Ökologie" (GfÖ) and the large community it represents in the German-speaking countries is a strategically important target and should be included in the list of priorities. With these changes, the current priority list is as follows:

- The Global Biodiversity Information Facility - GBIF
- Bundesamt für Naturschutz - BfN
- Research Data Alliance - RDA
- European Open Science Cloud - EOSC
- Bioschemas.org
- Gesellschaft für Ökologie - GfÖ
- German Network for Bioinformatics Infrastructure - de.NBI
- BMBF-Forschungsinitiative zum Erhalt der Artenvielfalt - FEdA

- Deutsches Zentrum für integrative Biodiversitätsforschung - iDiv
- Biodiversity Information Standards - TDWG

In the following section we will explain the measures with which we have begun to establish cooperations with various infrastructures and which measures are still planned.

2 Measures to strengthen NFDI4Biodiversity involvement

2.1 The Global Biodiversity Information Facility - GBIF

GBIF (https://www.gbif.org/) is a prominent player in the international biodiversity data landscape through its robust data infrastructure for global observation and specimen (occurrence) data as well as its checklist infrastructure. NFDI4Biodiversity partners are already engaged with GBIF on a variety of levels. The GBIF.de node system (see https://www.gbif.org/country/DE/summary) is run by the staff of the major German natural science collections and linked with NFDI4Biodiversity through data pipelines of the GFBio collection data centers. The GBIF-GFBio agreed pipelines for occurrence data are documented under http://gfbio.biowikifarm.net/wiki/Publication_of_Type_1_Data_via_BioCASe_Data_Pi pelines_at_GFBio_Data_Centers. The GBIF-D Node Manager (Triebel), the GBIF-D Speakers (Triebel, Grobe), several Representatives in the GBIF Governing Board (Häuser, Gemeinholzer, Güntsch), and several Node Staff Members are NFDI4Biodiversity partners and co-workers. A number of technical contributions and training activities in GBIF are delivered by NFDI4Biodiversity partner organisations (standards, software, components, workshops, etc.). Vice versa, GBIF tools are part of the NFDI4Biodiversity service portfolio and the German GBIF node organisations act as service providers in NFDI4Biodiversity, providing practical support to the Community in implementing GBIF tools like BioCASE provider software and publishing datasets in GBIF.

An additional important activity is now the development of the NFDI4Biodiversity use case "LAND" (Lebendiger Atlas Natur Deutschland) based on the GBIF "Hosted Portal" Platform (https://dev.gbif.org/hosted-portals.html). GBIF thus becomes an important component of the NFDI4Biodiversity infrastructure, which in turn provides data for open access via the GBIF Application Programming Interface (API).

2.2 Bundesamt für Naturschutz - BfN

The Federal Agency for Nature Conservation (BfN, https://www.bfn.de/en) is of outstanding importance in countless topics in the fields of biodiversity monitoring and protection. Consequently, NFDI4Biodiversity consortium partners are engaged in a variety of projects and cooperations with the BfN.

In NFDI4Biodiversity, a formal agreement was reached with the BfN that regulates access to the BfN's new checklist infrastructure via a designated API layer[1]. The agreement can be used as a blueprint for further cooperations with other infrastructures. In the future, it is planned to integrate the national checklist infrastructure more deeply internationally. Cooperation at the working level with the EU project TETTRIs (https://tettris.eu/) has been initiated for this purpose.

Of particular importance to this cooperation is also the National Monitoring Center for Biodiversity (NMZB, https://www.monitoringzentrum.de/) which was installed

---

[1] Referenz - https://checklisten.rotelistezentrum.de/api/public/swagger-ui

at the BfN in 2021. There have been several bilateral meetings between NFDI4Biodiversity and the NMZB, both on a strategic and working level, as well as participation in wider events such as the NMZB Forum or political meetups. The speaker of NFDI4Biodiversity has been appointed to the Advisory Group "Grundsatzfachgremium", and the NMZB Manager is a member of the NFDI4Biodiversity strategic advisory Board. NFDI4Biodiversity representatives are included in additional working groups (Fachgremien) for Data and Data management. From 2023 on, the teams will meet twice a year in an informal "jour fixe" to exchange on developments and mobilise synergies at the working level.

2.3 Research Data Alliance - RDA

The Research Data Alliance (RDA) is an international organisation that aims to develop and adopt infrastructure that promotes data-sharing and data-driven research (https://www.rd-alliance.org/). The RDA has over 10000 members from 145 countries.

RDA working and interest groups cover a huge range of topics in the field of data management in general, dealing for example with PIDs, metadata, or data management plans. The Biodiversity Data Integration Interest Group (https://www.rd-alliance.org/groups/biodiversity-data-integration-ig.html) focuses specifically on biodiversity data and in particular on semantic integration of biodiversity data. Also, the I-ADOPT framework is of interest[2].

Within the NFDI4Biodiversity partner organisations, there are currently numerous participations in various groups of the RDA, which, however, have usually emerged independently of the NFDI process. For the coming project phase, it is planned to match the existing participations, the range of interest groups, and the needs from the perspective of NFDI4Biodiversity. The goal is a strategic and targeted participation in the relevant working groups.

2.4 European Open Science Cloud - EOSC

The European Open Science Cloud (EOSC) aims to develop a Web of FAIR Data and services for science in Europe upon which a wide range of value-added services can be built. These range from visualisation and analytics to long-term information preservation or monitoring of the uptake of open science practices (https://eosc-portal.eu/about/eosc). The aspiration to structure the service landscape according to FAIR principles is obviously congruent with the vision of the NFDI4Biodiversity consortium. It is therefore a key concern to embed NFDI4Biodiversity services into the evolving EOSC landscape and to achieve the best possible interoperability and performance.

Various measures were taken to initiate interaction with the EOSC:

On 1. April 2022 Wolfram Horstmann from the EOSC Executive Committee participated in a TA2 ("2connect") video call to explain the aims and structure of the EOSC, its connection with various NFDIs in general, and to enter into a discussion process on interaction with NFDI4Biodiversity. In particular, the possibility of including NFDI services in the EOSC Marketplace was seen as an appropriate first step. The documentation of services from the NFDI4Biodiversity consortium currently being developed in WP3 provides an excellent basis for registering services in the EOSC Marketplace (and potentially other systems) in the

---

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

future. It is planned to evaluate the services documented by then in spring 2024 and, based on this assessment, to carry out pilot registrations in the Marketplace.

To date, three NFDI4Biodiversity partners have registered in the Marketplace providing four concrete services: SSH Open Marketplace and B2Share (Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG)), de.NBI Cloud (de.NBI), and BacDive (German Collection of Microorganisms and Cell Cultures (DSMZ)). The Alfred Wegener Institute (AWI), the Center for Marine Environmental Sciences, the University of Bremen (MARUM), and the Heidelberg Institute for Theoretical Studies (HITS) have registered to the Marketplace but do not offer dedicated services yet.

2.5 Bioschemas.org

At the NFDI4Biodiversity All Hands Conference 2022 Franck Michel from bioschemas.org gave a presentation about Schema.org and its bioschemas.org extensions. In the following discussion it was decided to make Schema.org with the bioschemas.org extensions the recommended data format to deliver dataset metadata information from the various data providers and components of NFDI4Biodiversity to the search harvester. Since various components are involved with the usage of bioschemas.org within the consortium, two workshops will be held to coordinate and harmonise its usage with representatives from all of the major technical components. The first workshop will take place in May 2023 and the second one is currently planned for the fall of 2023.

Additionally, we have been in contact with representatives from other NFDI consortia in regard to their usage of schema.org.

2.6 Gesellschaft für Ökologie - GfÖ

The Ecological Society for Germany, Austria, and Switzerland (GfÖ) is an independent, nonprofit scientific organisation that aims to promote basic and applied ecological science, encourage collaborative work of all ecological disciplines, improve communication among ecologists in German-speaking countries and beyond, and facilitate education in ecology at universities and institutes of higher education (https://gfoe.org/en). Due to the focus on ecological data, the large number of members, and the focus on German-speaking countries, the GfÖ represents an extremely important user community for the data services developed in NFDI4Biodiversity.

NFDI4Biodiversity consortium partners already have many links to the GfÖ, and the GfÖ is a participant in the NFDI4Biodiversity consortium. In this capacity, a joint five-day Winter School was organised in December 2022, which was very well received (about 140 applications for 20 seats). For several years, the GFBio/NFDI4Biodiversity Research Data Management workshops have been a continuous element in the pre-conference programme of the Annual GfÖ Conference (Jahrestagung), with 30-40 participants. In order to make the infrastructure and service portfolio even more widely known, a session on National Data Infrastructures has been submitted and accepted in the main programme of the GfÖ Conference 2023. Poster contributions and an information booth serve to disseminate information about the NFDI services among conference participants. Further engagement until 2025 and potential activities in a next funding phase will be discussed with the GfÖ Board in Summer 2023.

2.7 German Network for Bioinformatics Infrastructure - de.NBI

The German Network for Bioinformatics Infrastructure (de.NBI) was originally launched and funded by the German Federal Ministry of Education and Research (BMBF) to sustainably maintain important bioinformatics infrastructures such as software and cloud computing for life science and to make them more accessible. Although the BMBF has discontinued funding for the network at the end of 2021, federal funds have been made available to enable the continuation of de.NBI. In order to continue to provide a common platform for the scientists involved or interested in de.NBI, the non-profit association de.NBI e.V. was founded in 2021.

As of 2022 de.NBI e.V. is a participant in NFDI4Biodiversity with the following tasks:

- Secure access to the de.NBI Cloud
- Support and advise in the context of multi-cloud capable services, distributed data storage, and function shipping
- Cloud-related training

The membership of de.NBI e.V. complements the activities of de.NBI service centers Gießen and Bielefeld, which are Co-Applicants in NFDI4Biodiversity. The de.NBI Cloud is an important building block in the NFDI4Biodiversity IT infrastructure. Major components of the Research Data Commons architecture are hosted in the de.NBI Cloud (Core Storage, Data Discovery Portal) and two Use Case projects have migrated their data analysis pipelines to the use of the de.NBI Cloud.

2.8 BMBF-Forschungsinitiative zum Erhalt der Artenvielfalt - FEdA

FEdA is a BMBF research initiative that addresses the scientific basis of biodiversity and concrete strategies against the loss of biodiversity. Comprehensive research data on changes in biodiversity and the effectiveness of various measures are being generated in the FEdA projects now and in the coming years. These data sets are to be made accessible for further scientific analyses and made findable in a central portal.

To do so, the data must be archived and made accessible in a professional manner. To this end, FEdA and NFDI4Biodiversity agreed on cooperation in the following areas, among others (https://www.nfdi4biodiversity.org/de/kooperation-feda/):

- the transfer of knowledge on the subject of research data management
- advising and supporting the funded projects ("FEdA projects") in the preparation of data management plans
- supporting FEdA projects in archiving research data in suitable data centers, including the necessary links to each other and to the project ("distributed archiving")
- the development of a process for the delivery of metadata from the NFDI4Biodiversity consortium data centers to the FEdA coordinating body for use in the planned portal.

Since the cooperation with FEdA began, eight FEdA projects have received support on RDM through exchange with the GFBio's DMP officer. To this day, there are currently four interdisciplinary FEdA projects receiving dedicated support in the development of a data management plan (DMP), the number is expected to increase in the upcoming months.

Three training workshops in RDM, DMP development, and data submission and archival have been provided to the FEdA researchers. More details about the

cooperation between FEdA and NFDI4Biodiversity are available here: https://zenodo.org/record/7624583#.ZDlFinZBy3B.

2.9 Deutsches Zentrum für integrative Biodiversitätsforschung - iDiv

The German Centre for Integrative Biodiversity Research (iDiv) is a DFG research center where researchers from 40 nations establish the scientific basis for the sustainable management of earth's biodiversity, including not only species diversity but also genetic diversity, diversity of functions, interactions, and ecosystems. Several iDiv researchers are Co-Applicants in NFDI4Biodiversity (e.g. König-Ries and Bonn) and are shaping the work within the different Task Areas. Here, the iDiv is especially involved in Community Engagement with a focus on managing several of the 26 NFDI4Biodiversity Use Cases and on running the Use Case Plant Hub (https://planthub.idiv.de/).

2.10 Biodiversity Information Standards - TDWG

TDWG Biodiversity Information Standards (https://www.tdwg.org/) is a global standardisation organisation in the field of data-driven biodiversity research. TDWG standards relate to a wide range of aspects, such as data formats, access protocols, API specifications, but also entire data citation frameworks. A large number of NFDI4Biodiversity partners are involved in TDWG activities and contribute to the specification and ratification of standards.

In addition to existing activities, NFDI4Biodiversity seeks to put the topic of National Data Infrastructures on the agenda as an important contribution to international efforts in the field of biodiversity informatics. With this aim, a symposium "National biodiversity data centers: challenges and opportunities" was organised in 2022 at the TDWG conference in Sofia, Bulgaria[3]. In that symposium, a talk about NFDI4Biodiversity was presented [2]. Based on the results of the symposium and coordinated by NFDI4Biodiversity, a publication is currently being prepared that addresses the need for national data infrastructures and their delimitation from international services.

3 Outlook

The priorities and respective actions for collaborations with external initiatives, services, and infrastructures described in this report represent a snapshot of the embedding of NFDI4Biodiversity in the broader landscape of the biodiversity community. Priorities will continuously be adapted in the future and the appropriate instruments for cooperation may change as tasks shift or new infrastructure components become available. One new prospect for collaboration will be the focus group Biodiversity of the ELIXIR network[4], which will launch a Community Implementation Study in June 2023. Part of the Implementation Study will be outreach and networking activities which invite initiatives such as NFDI4Biodiversity to connect with the emerging ELIXIR biodiversity community. The statuses of the cooperations are continuously being documented in the collaborative working environment of the consortium (Confluence). Similarly, the results of discussion processes, e.g. in the Special Interest Group "Strategy" or in Task Area 2 ("2connect"), can always be viewed and updated transparently in this dynamic documentation system.
In addition, we will place a stronger focus on concrete measures, e.g. in the area of standardisation, since with the progressive development of the Research Data Commons reference implementation (RDC) and the use cases based on it, it is to

---

[3] https://biss.pensoft.net/collection/389/
[4] https://elixir-europe.org/focus-groups/biodiversity

be expected that new requirements for the availability and integrability of external resources will arise.

4 References

[1] Güntsch, A., Weber, J., Ebert, B., Frenzel, M. (2021): NFDI4Biodiversity - First strategic report on national and international involvement.
https://zenodo.org/record/8297699

 [2] Luther, K., Güntsch, A., König-Ries, B., Fichtmueller, D. (2022): NFDI4Biodiversity: a German infrastructure for biodiversity data. Biodiversity Information Science and Standards 6: e93869. https://doi.org/10.3897/biss.6.93869