Linking Domain-specific RDM to Institutional and Generic Approaches – the Case of NFDI4Biodiversity

Jimena Linares[1], Barbara Ebert[1], Judith Sophie Engel[2,3]

[1]German Federation for Biological Data – GFBio e.V.;
[2]University of Bremen;
[3]Center for Environmental Sciences - MARUM

NFDI4Biodiversity is a consortium within the German National Research Data Infrastructure – NFDI. The consortium has been in action since October 2021 with the main goal of mobilizing data, rolling out services, and offering tools for the biodiversity community. At the moment, it has over 50 partners and a five-year work program until 2025, funded through a joint initiative of the Federal Government and its states (the Länder).

One of the partners is the German Federation for Biological Data – GFBio (e.V). GFBio has been assisting scientists and data managers for years regarding all types of inquiries about research data management in biodiversity, ecological, and environmental research, and providing a portfolio of services along the research data life cycle.

With an expanding consortium and increasing interest in research data management in universities and research-performing organizations, one challenge to overcome is how to support the scientific community. On-site (e.g. on-campus) interaction with researchers is often provided by local research data management services (RDM officers). Still, it is common that resources are limited and in-depth subject-specific advice cannot be provided. The "Front Office/Back Office" model propagated by GFBio presents a good way to serve the scientific community by linking domain-specific approaches with institutional and generic ones.

1 Introduction

Community-specific tools and expertise are certainly needed for better management of research data. NFDI4Biodiversity[1] works with different data and service providers in the domain of biodiversity, ecology, and environmental research (Weber et al. 2021). The consortium is composed of and supports a diverse community of researchers, research

Publiziert in: Vincent Heuveline, Nina Bisheh und Philipp Kling (Hg.): E-Science-Tage 2023. Empower Your Research – Preserve Your Data. Heidelberg: heiBOOKS, 2023. DOI: https://doi.org/10.11588/heibooks.1288.c18062

1 https://www.nfdi4biodiversity.org/en/; *Last accessed on March 7th, 2023.*

organizations, public authorities, professional societies, and citizen science initiatives in managing their data and making them available for re-use. The community mainly generates data from field and lab work. Examples of these data are: occurrence data (observations of species); environmental data (e.g. temperature, rainfall); trait data (e.g. seed number and mass); molecular data (e.g. DNA and RNA sequences); experimental and laboratory measurements; multimedia (photographs, audio, and video of e.g., orthophotos produced using a drone, observed specimens); digital surface models; model code and statistics. Publication and distributed archiving are organized in a network of ten specialist data centers coordinated by the German Federation for Biological Data (GFBio[2]). Seven data centers are located at natural sciences collections across Germany. There, researchers can submit and get advice about the best way to submit their data according to their affiliations, specific data type (e.g. audio from birds callings), archival, and publication intentions. The other three data centers are dedicated to handling nucleotide, plant, and environmental data. The consortium's Helpdesk is the central entry point for requests from partners and researchers related to their research data management activities, including the submission of data for archiving and publication. The Helpdesk is a service run and maintained by the consortium partner GFBio e.V.[3]and is set up in the Jira project manager software[4]. The ticket workflow in the Helpdesk is a centralized dispatching system. Requests are assigned to a team of experts within NFDI4Biodiversity and cooperating partners by the Helpdesk Team.

2 Core services for biodiversity, ecological, and environmental research

The Helpdesk acts as the gateway for all requests. User requests via mail to helpdesk@nfdi4biodiversity.org and info@gfbio.org or via the contact form on nfdi4biodiversity.org generate a ticket in the Helpdesk. The Data Management Planning Tool - DMPT and the data submission service (which will be discussed later) are also connected to the Helpdesk. There are three major request types (Figure 1): 1. HELP requests for general requests around (research) data management, service support, training, event, or networking; 2. DMP-type requests (a specialized type of HELP request) generated by users that use the DMPT and require personalized support, and 3. DSUB requests generated when users are submitting data via the data submission service (aka submission system).

All communication takes place through the Helpdesk in the respective ticket(Astor et al. 2021), so the information is documented there in a transparent manner for the users, and it is internally reachable for colleagues involved and interested in the ticket. The Helpdesk workflow promotes the Front Office/Back Office model (Figure 2): The Back Office comprises the Helpdesk Team (first-level support) and the expert network in the consortium (second-level support). The Front Office is the expert network at the univer-

2 `https://www.gfbio.org/data-centers`; *Last accessed on March 22nd, 2023.*
3 `https://www.gfbio.org/gfbio_ev/`; *Last accessed on March 10th, 2023.*
4 `https://www.atlassian.com/software/jira`; *Last accessed on March 7th, 2023.*

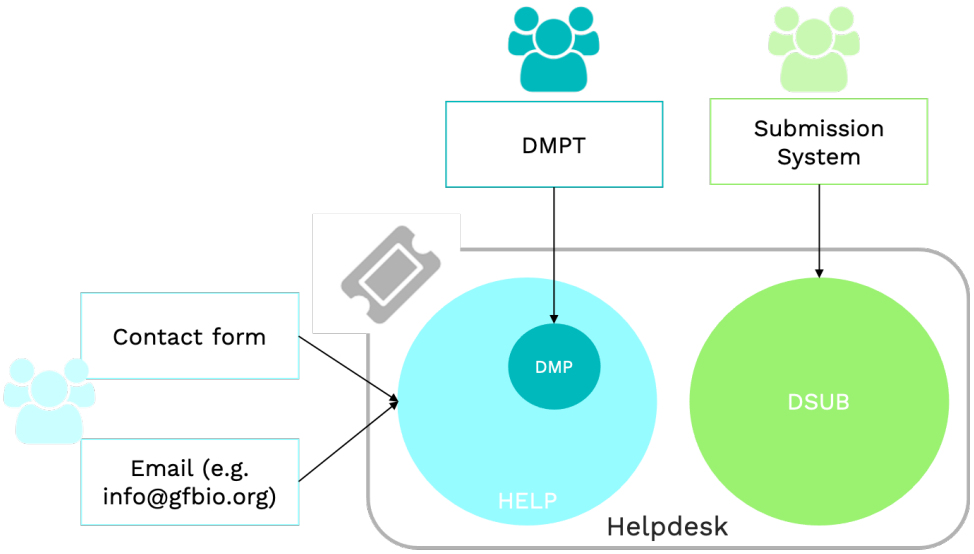

Figure 1: Helpdesk requests, including two core services in RDM.

sities and research institutions, e.g. the data stewards or data managers that offer on-site support to students and scientists of their organization.

They can address discipline-specific questions to the Helpdesk Team, or forward the specific questions directly to the Helpdesk Back Office.

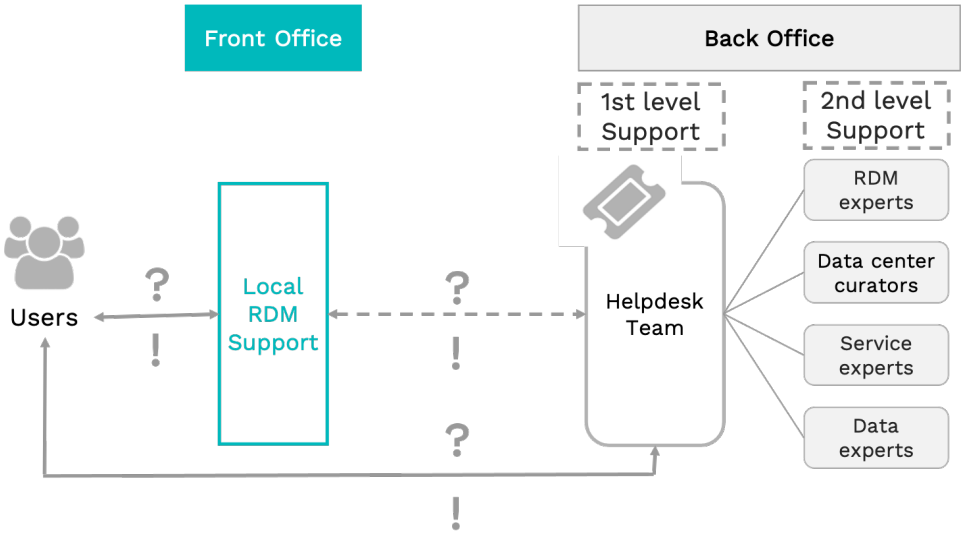

Figure 2: Front Office/Back Office support provided by the Helpdesk.

The Front Office/Back Office model has been continuously tested in two pilot projects:

a) FEdA & KonsortSWD The "BMBF Research Initiative for the Conservation of Biodiversity" (in German: Forschungsinitiative zum Erhalt der Artenvielfalt) – FEdA[5]

5 `https://www.feda.bio/en/`; *Last accessed on March 20th, 2023.*

is a long-term funding program with several cohorts of interdisciplinary projects on biodiversity and biodiversity protection. The funder expects the projects to deliver a data management plan and metadata as part of their data outputs, to feed a joint data portal for the program. The speakers of FEdA and NFDI4Biodiversity agreed to collaborate early on to link the research initiative's data management efforts to the evolving support structure in the NFDI. A focal point of the collaboration is the support of data management plans for these projects. The data manager from FEdA is included in the NFDI4Biodiversity Helpdesk, and a workflow was established to coordinate the data management plan and data archival support. The workflow has been optimized such that FEdA assumes the Front Office role, providing the initial support and dissemination of information. Then, requests are forwarded to the respective expert network in the "Back Office" of NFDI4Biodiversity for specific biological, ecological, and environmental research questions.

As FEdA's interdisciplinary projects often have social sciences components, a cooperation with KonsortSWD[6], the NFDI Consortium for social, behavioral, educational, and economic science, and specifically Qualiservice[7], was established to include additional support by social data experts. In this way, the collaboration with FEdA is also a use case for cross-consortia data management support. The Helpdesk workflows were adapted so the data experts can communicate with the researchers and the NFDI4Biodiversity colleagues as well as incorporate their input in the data management plans. When researchers decide to archive their interview data with Qualiservice, support for the necessary documentation and publication procedures are moved to the Qualiservice helpdesk, which is easier for both sides. An example is the use case project "SLInBio - Urban lifestyles and the valorization of biodiversity: dragonflies, grasshoppers, bumblebees and Co", which researches how the perception and valorization of insects can be increased in an area and what contribution cities can make to the conservation of insect diversity[8] The project works, among other data types with molecular and ecotoxicological data, interviews, and surveys. Each of them will be archived and published in the most fitted repository, which implies that diverse repositories for different disciplines (biological, environmental, and social sciences) will be used for publication.

b) UBRA/ Data Train The Helpdesk collaborates with the U Bremen Research Alliance – UBRA[9], specifically with the coordinator of the Research Data Management and Data Science Training course "Data Train"[10] for planning and organizing the courses on research data management. This serves as a notable instance of continuous collaboration in research data management with state initiatives (Landesinitiativen) and institutions, such as the University of Bremen.

6 `https://www.konsortswd.de/en/`; *Last accessed on March 19th, 2023.*
7 `https://www.qualiservice.org/en/`; *Last accessed on March 19th, 2023.*
8 `https://www.isoe.de/en/nc/research/projects/project/slinbio-1/`; *Last accessed on July 31st, 2023.*
9 `https://www.bremen-research.de/en/`; *Last accessed on March 22nd, 2023.*
10 `https://www.bremen-research.de/data-train/`; *Last accessed on March 22nd, 2023.*

As mentioned, the Helpdesk tickets also involve two special types of tickets that are produced once the users access and make use of two core services: data management plan support (through the DMPT) and the data submission service. These two are the most popular services offered by the consortium and, together with other services[11], aim to facilitate the best practices in research data management following the FAIR principles (Wilkinson et al. 2016), as well as domain-specific data standards and practices. Both services are explained below:

Data management plan support

The DMPT[12] has been the go-to tool for supporting and conceptualizing Data Management Plans – DMPs in biodiversity research for years, with a special focus on DFG requirements. In 2021, the DMPT underwent a re-implementation using the Research Data Management Organizer – RDMO (Klar et al. 2023) as a backbone, while retaining its familiar front end. This was done to optimize personalized DMPs for specific cases and to facilitate interdisciplinary cooperation, among other benefits. The revamped DMPT-RDMO base is an example of a specific RDM tool transitioning from an in-house development to a community-supported open-source tool, ensuring both sustainability and better support of research data management efforts across disciplines and institutions. The tool's questionnaire includes the common components of a DMP, and its sections follow the Guidelines of the German Research Foundation (Deutsche Forschungsgemeinschaft) – DFG[13] on Handling of Research Data[14], Handling of Research Data in Biodiversity Research[15], and for Safeguarding Good Research Practice[16]. The support of individual researchers on the making of a DMP as well as on the posterior archival and publication of the data that will be generated in the research project is a collaborative effort among experts who work with domain-specific data types. The DMP produced follows consensus among all involved parties (as defined by their roles: user, DMP Officer, DMP controller, collaborators such as contacts from data centers involved), and other experts (e.g. colleagues from initiatives such as FEdA, colleagues from other consortia). An example of a GFBio-approved (consented) DMP can be seen in the model DMP (Mau, Timmermann, and Astor 2020).

The data submission service

The data submission service[17] works with the above-mentioned data centers to offer a comprehensive curation service of the research data provided to GFBio (as seen in Figure 1).

11 `https://www.gfbio.org/services`; *Last accessed on March 29th, 2023.*
12 `http://dmp.gfbio.org`; *Last accessed on March 29th, 2023.*
13 `https://www.dfg.de/`; *Last accessed on March 29th, 2023.*
14 `https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/forschungsdaten/guidelines_research_data.pdf`; *Last accessed on March 29th, 2023.*
15 `https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/forschungsdaten/guidelines_biodiversity_research.pdf`; *Last accessed on March 29th, 2023.*
16 `GuidelinesforSafeguardingGoodResearchPractice`; *Last accessed on March 24th, 2023.*
17 `http://submissions.gfbio.org`; *Last accessed on March 20th, 2023.*

The data centers include international data archives such as the Data Publisher – PANGAEA[18] or the European Nucleotide Archive – ENA[19] as well as several specialized German Collection Data Centers. Services are organized along five major biological data types[20]. All data centers have agreed on common (meta)data standards[21] for submission and curation. There is a minimum of consensus metadata required before uploading the data. Data center staff cooperates through the Helpdesk infrastructure to organize distributed archiving. In this way, thematically and structurally heterogeneous data can be archived in different data centers (e.g. ENA and PANGAEA), and the persistent identifiers will be linked with each other. The funding obtained through the National Research Data Infrastructure is crucial to ensure this engagement.

To create a submission through GFBio, the user is required to obtain a GFBio account to access the submission service. This can be done with the individual Academic ID, or by obtaining an ID from the GFBio Identity provider service. After that, the direct support on the specific submission and curation begins. Each submission requires individual support by a Helpdesk Team member, as there can be cases where the data files and/or types do not fit within the submission criteria of the data centers, so external alternatives for archival and publication are suggested to the submitters. File sizes in the submission helpdesk are currently limited to 200 MB. This is different from the size of files accepted in each data center for archival and publication (ENA, for example, has no restrictions). The curation of datasets is supported by the curators of the data centers through communication with the submitter in the ticket, which is automatically created by the submission request made through the helpdesk. For general requests about submissions landing on the helpdesk, besides the support provided by the experts in the request, there is also a self-serve online library, the Knowledge Base[22], with information for current and future submitters[23]. Depending on the type of data, dedicated templates, and documentation are also available to facilitate standardization. For example, the molecular submission template[24] for datasets aimed to be brokered to ENA has documentation on the formats and values expected[25] for the minimum mandatory and optional fields. This documentation is continuously updated by the GFBio submissions team to comply with the latest requirements of ENA. The finished dataset is archived in the corresponding data center with a persistent identifier (usually DOI for non-molecular or an accession number for molecular data) for a long time.

18 `https://www.pangaea.de/`; *Last accessed on March 29th, 2023.*

19 `https://www.ebi.ac.uk/ena`; *Last accessed on April 4th, 2023.*

20 `https://gfbio.biowikifarm.net/wiki/Major_Types_of_Biological_Data`; *Last accessed on July 31st, 2023.*

21 `https://gfbio.biowikifarm.net/wiki/Data_exchange_standards,_protocols_and_formats_relevant_for_the_collection_data_domain_within_the_GFBio_network`; *Last accessed on July 31st, 2023.*

22 `https://kb.gfbio.org/display/KB/Knowledge+Base`; *Last accessed on July 31st, 2023.*

23 `https://kb.gfbio.org/dosearchsite.action?queryString=submission&where=KB&additional=page+excerpt&labels=faq&contentType=page`; *Last accessed on July 31st, 2023.*

24 `https://gitlab.gwdg.de/gfbio/molecular-submission-templates/-/raw/master/full_template.csv?inline=false`; *Last accessed on March 31st, 2023.*

25 `https://gitlab-pe.gwdg.de/gfbio/molecular-submission-templates/-/blob/master/Template-Description.md`; *Last accessed on March 31st, 2023.*

3 Beyond the integration of the services in a single domain - the Helpdesk as the focal point

Feedback shows that the community appreciates the support provided by the Helpdesk and the human component of the assistance (as it is possible to address a contact person as an "assignee" on the ticket) during and after the consultation. Indicators of success for the team behind the Helpdesk are the loyalty of many users who have relied on GFBio for years, the dissemination among the scientific community evidenced in new requests submitted to the Helpdesk, the growing number of cooperations with projects and institutions, and the gratitude of users once their consultation is completed.

Working on projects such as the pilot with FEdA & KonsortSWD is a good example of the usability, potential, and expansion of the Helpdesk's model to also integrate interdisciplinary exchange in one single service point. In the case of FEdA, both the support in the creation and completion of DMPs as well as support through the Helpdesk has led to an extended cooperation formalized as a 'memorandum of understanding' between the speakers of FEdA and the NFDI4Biodiversity consortium for continuous and future support on the development of (interdisciplinary) data management support[26].

Naturally, there have been challenges when working in interdisciplinary RDM support. The social sciences have carefully designed workflows to guide their users through a multi-step process for archiving person-related research data early on in the research process, including support with informed consent, formalized study reports, and data transfer agreements. These are different from the archival support for biological data types, which is mostly scheduled toward the end of the research process and based on a structured DMP. Although the difference in expectations regarding the consultation process was initially challenging, we found the early engagement with archiving services a useful concept to improve relations with our core community. There have been mutual learning benefits throughout the process of encouraging researchers in their effort towards reaching FAIRness and producing consented interdisciplinary DMPs.Working together on data management planning has made us re-think the ways of incorporating RDM practices inside the scientific research process. We are taking action by providing workshops, having 1-to-1 support encounters with the researchers, and creating self-learning materials.

The NFDI4Biodiversity Helpdesk is also committed to providing further support beyond the consortium's offers. If there are inquiries in RDM (including DMPs) in another domain that reach the Helpdesk, the team will suggest or direct users to an appropriate point of support, such as the local RDM officer or experts in the respective field. The latter also implies communication with other NFDI Consortia. Accordingly, if certain submitted data do not thematically or qualitatively belong to any of the GFBio data centers, alternatives for archiving and publication are suggested to the submitter.

26 `https://www.nfdi4biodiversity.org/en/who-we-are/cooperation-feda/`; *Last accessed on April 3rd, 2023.*

4 Conclusions

With more than 100 requests processed since the NFDI4Biodiversity consortium began, its Helpdesk has proven successful as a central entry point of access to the consortium's services, tools, and expertise. Furthermore, the Helpdesk is also facilitating transparent communication among the community working in biodiversity, ecological, and environmental research.

This is also evidenced by the Front Office/Back Office. The pilot cooperations with the BMBF Research Initiative for the Conservation of Biodiversity (FEdA) and the UBRA Data Train are examples of how to build productive working relations with institutions, universities, and research centers, e.g. by allowing local RDM officers to create and address support tickets and to participate directly in the consultations together with domain experts from the consortium. Within NFDI4Biodiversity, we aim to encourage more universities and institutions to implement this model in their RDM support on-site and to expand the number of Front Offices step by step.

The pilot cooperation with FEdA is an example of how interdisciplinary research projects can be effectively supported in the National Research Data Infrastructure NFDI. As this cooperation continues and new project cohorts start, we will need to enlarge the group of partners (e.g. from the medical domain) due to the new disciplines involved.

We think that over time Helpdesk services like ours could consolidate across consortia, especially where similar data types or research methods are concerned. This would support the transition to a "ONE NFDI" with interoperable, user-facing services across consortia. Linking domain-specific and local or generic support structures is an opportunity to create workflows that make it easy for any researcher to find the best service for their data management issues.

Acknowledgments

The GFBio project as well as the NFDI4Biodiversity consortium have been funded by the German Research Foundation -DFG under the grant agreement numbers 229241684[27] and 408180549[28] (GFBio) and 442032008[29] (NFDI4Biodiversity). NFDI4Biodiversity is part of NFDI, the National Research Data Infrastructure Program in Germany.

References

Astor, Tina, Judith Weber, Ivaylo Kostadinov, Frank Oliver Glöckner, and Jens Nieschulze. 2021. "Potential für ein starkes Netzwerk zwischen GFBio und FDM-Bera-

27 `https://gepris.dfg.de/gepris/projekt/229241684`; *Last accessed on May 8th, 2023.*
28 `https://gepris.dfg.de/gepris/projekt/408180549`; *Last accessed on May 8th, 2023.*
29 `https://gepris.dfg.de/gepris/projekt/442032008?context=projekt&task=showDetail&id=442032008&`; *Last accessed on May 8th, 2023.*

tenden an Universitäten und Forschungsinstituten". *Bausteine Forschungsdatenmanagement,* number 1: 22–31. DOI: https://doi.org/10.17192/BFDM.2021.1.8311.

Klar, Jochen, Olaf Michaelis, David Wallace, Max Schröder, Enke Harry, Giacomo Lanza, David Martínez Muñoz, and Dario Pilori. 2023. *Research Data Management Organizer (RDMO).* DOI: https://doi.org/10.5281/ZENODO.596581.

Mau, Franziska, Britta Timmermann, and Tina Astor. 2020. *GFBio Model Data Management Plan (DMP).* DOI: https://doi.org/10.25625/W3YEEQ.

Weber, Judith, Barbara Ebert, Michael Diepenbroek, Ivaylo Kostadinov, and Frank Oliver Glöckner. 2021. "NFDI4BioDiversity – NFDI-Konsortium für Biodiversitäts-, Ökologische und Umweltdaten". *Bausteine Forschungsdatenmanagement,* number 2: 98–109. DOI: https://doi.org/10.17192/bfdm.2021.2.8334.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersbergand, Gabrielle Appleton, Myles Axtonand, Arie Baakand, Niklas Blombergand, et al. 2016. "The FAIR Guiding Principles for scientific data management and stewardship". *Scientific data* 3 (1): 1–9. DOI: https://doi.org/10.1038/sdata.2016.18.