Memorandum of Understanding (v2)

The speakers of the consortia DataPLANT, NFDI4Microbiota, NFDI4Biodiversity, FAIRagro, NFDI4Objects, and NFDI4BIOIMAGE agree on the following areas of cooperation in the area of research data management for the biological sciences.

About the consortia

DataPLANT is a consortium for the plant science community, providing a state-of-the-art tool-box for managing, processing and publishing data packages early on in the research process following an open contribution and participatory model. Its overarching goal is to provide robust, easily accessible and user-oriented research data management practices, tools, and infrastructure to support collaborative plant biology research.

NFDI4Microbiota offers a broad set of solutions and services to the microbiological research community to make their data and other research output FAIR and open. This includes among others training, data storage solutions, data analysis services and RDM consulting.

NFDI4Biodiversity is a broad resource provider and expert network with a variety of services for biodiversity, ecological and environmental data. In the first funding phase, the foundations were laid for a common infrastructure and data mobilisation according to the FAIR principles, with focus on occurrence and molecular sequencing data (metabarcoding).

FAIRagro is developing a research data infrastructure for agrosystem sciences. Quality-assured and legally compliant datasets are connected and made discoverable, reusable and machine-actionable in accordance with FAIR principles. Support for data analysis and research data management is provided through a coordinated network of trainers, data stewards, and a central helpdesk.

NFDI4Objects develops solutions for research data management related to the material heritage of approximately three million years of human and environmental history. The consortium is aimed at individuals of all experience and career levels who wish to collaboratively shape the digital transformation of research data infrastructure within the involved disciplines in a sustainable manner.

Image data is central to biological research because it captures spatial, temporal, and functional relationships in high resolution, e.g., in disease research. **NFDI4BIOIMAGE** develops new formats, metadata standards, analysis methods, and processing environments for optimal use, e.g., in AI-supported pattern recognition.

Goals and areas of cooperation

Despite different subject areas and research contexts (field/lab), the target communities of the six consortia use similar methods and similar data types are relevant for their research. Based on their relevance for the community, the consortia's investments in solutions for certain data types and methods in the first three years of operation were quite complementary. This provides us with a solid basis to leverage resources.

Table 1: Complementary investment in major biological data types
The focus of investment of DataPLANT, NFDI4Microbiota, NFDI4Biodiversity, FAIRagro, NFDI4Objects, and NFDI4BIOIMAGE in community services, guidelines, blueprints, use cases, training, etc. for different biological data types were highly complementary in the first three years of operation.

Data type	DataPLANT	Microbiota	Biodiversity	FAIRagro	Objects	BIOIMAGE
Type 1: Collection data, species occurrence data	-	-	+++	+	+++	-
Type 2: Taxon data (catalogue, checklist or so-called red list)	-	+	+++	+	+	-
Type 3: biological and ecological study data including functional and phylogenetic trait data	+	++	+	+++	+	++
Type 4: non-molecular analysis data f.e. in laboratory research contexts	+++	+++	+	++	++	+++
Type 5: molecular (sequence) data	+++	+++	++	+++	+	+

***Legend**: +++ high; ++ solid; + not much in focus; - no priority so far*

We therefore intend to use biological data, along with related software, tools, workflows, models, and other digital infrastructure, as the connecting element to harmonise research data management strategies for different research contexts, thereby facilitating access to services and other data-related resources in the field of biology, and developing synergies across consortia. The cooperation will be open to include other consortia with biological data.

Synergies may include but are not restricted to:

- Education and training
- Outreach and communication
- Service portfolio (reuse and co-development of services)
- User support
- Activities in the UN policy arena regarding digital sequence information

Living the collaboration: Activities

The consortia will implement a biodata interest group within NFDI to govern joint activities and to include other consortia over time. The interest group shall review training materials and service developments to highlight synergies and complementarity, so that operational goals in the consortia can be aligned and co-marketing activities can be planned. It will develop cross-consortia collaborations on essential services and advocate for sustained funding of biodata services. The interest group shall be integrated in the work programmes and coordination shall alternate between the consortia.

1) **Joint partners and staff**: As all involved consortia have started or are approaching the second funding period, they will strengthen their partnership through shared co-applicants, participants and staff. A list of current partners is included as an annex to this MoU. Shared partners will allow the co-funding of staff positions for core activities in service development, marketing, and in the policy sector.

2) **Joint Helpdesk and User Support:** The consortia will coordinate their user support activities by aligning helpdesk structures and establishing shared triage and referral processes. This collaboration will be developed in close connection with the planned Base4NFDI proposal Support4RDM, which explores the networking of helpdesks across the NFDI.

3) **Joint infrastructure**: The consortia rely on collaboration with service centres and cloud such as provided by de.NBI. They have a common interest to stimulate the further development of such infrastructure to meet the needs of their target communities and will provide new use cases encouraging the sharpening of portfolios for research data infrastructure providers in the life sciences. This will be facilitated through the service and cloud providers who are shared members of the consortia. There is also a joint interest in multicloud technologies and the Research Data Commons concept. This will be promoted jointly within the current negotiations about the NFDI architecture and basic services.

   Additionally, as part of the joint infrastructure strategy, the consortia will explore the establishment of a shared data hub as a long-term, cross-consortia platform for collaborative data curation and publication. The institutions will determine and assume responsibilities for hosting, operation, and maintenance as the collaboration progresses.

4) **Joint programming for the uptake of services and standards**: The consortia will develop dedicated cross-consortia flexfunds projects and use cases, for example regarding the use of **FAIR Digital Objects**, e.g. ARC (the Annotated Research Context) or OME-Zarr; **workflows**, e.g., SciWin (Scientific Workflow Infrastructure), or CloWM (Cloud-based Workflow Manager); **benchmarking tools** (e.g., the CAMI benchmarking portal); **storage infrastructure** (e.g., Aruna); an

overarching search service; **ontology portal** (e.g.; BiodivPortal and Bioportal), and eventually TS4NFDI; the joint development of tools for the collection of service **usage metrics**; and the aligned implementation of **schemas** (e.g., bioschemas.org) by data and service providers in the consortia. These projects and use cases will be used to promote services across subject areas and to plan cross-consortia workshops and seasonal schools. For services marketed across consortia, joint user groups will be set up.

5) **Joint scientific conferences**: The consortia are organizing a joint conference. This will promote scientific exchange between the various consortia and involve the user communities. Possible main topics are data engineering, data management, interoperability concepts, data stewardship, data science, reference data sets and use cases. Such a conference should be organized alternately by at least one consortium and focus on the subject area of the respective consortium. This series and other meetings enable the exchange of ideas and discussion, allowing the consortia to refine their approaches, gain new perspectives, and build networks with the communities. They also inspire interdisciplinary learning and spark collaborations that lead to joint publications, project proposals, and data sharing initiatives.

6) **Joint publication activities:** The consortia will pursue coordinated publication activities, including the preparation of special issues in relevant peer-reviewed journals (e.g. Journal of Integrative Bioinformatics (JIB)) to disseminate outcomes of cross-consortia projects, conferences, and community initiatives.

7) **Pooling know-how for outreach purposes**: The consortia use branded knowledge bases to reach their target audiences and market their expertise and services. The consortia strive to pool capacities for the development and update of such resources, in order to achieve economies of scale and avoid duplication of efforts. Care will be taken to enable marketing in each consortium´s community of interest, while providing easy access to comprehensive biodata resources for biology in general. Collaboration with existing knowledge brokers in the field, e.g. ELIXIR, forschungsdaten.info, WiNoDa, and GFBio e.V., will be pursued as part of this activity.

8) **Joint training**: The consortia will harmonize their training offerings and will try to complement each other's training portfolio. They will conduct co-hosted training events and will share and co-create training material.

Transparency

The partners provide information on the collaboration on their websites.