Following the spirit of the "Berlin Declaration on NFDI Cross-Cutting Topics” discussed and published in August 2019, members of 14 consortia and 7 cross-cutting initiatives met again in February 2020 to discuss opportunities and challenges for a joint development of a national research data infrastructure. As a contribution to an open and joint discussion of Cross-Cutting Topics NFDI4BioDiversity decided to publish the chapter „The Consortium within the NFDI“.

The chapter is an excerpt of the NFDI4BioDiversity consortium proposal submitted on 15.10.2019 to the DFG for the first funding period. It is intended for an open, transparent inter-consortia discussion of cross-cutting topics.

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](image)

**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”.

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, 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, DINI e.V., nestor, or the new CoreTrustSeal have become widespread means. Based on our long-term experiences with the certification of the PANGAEA information system, our essential role in the development of the CoreTrustSeal, and our involvement in the H2020 FAIRsFAIR project, 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), 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.

Bremen, 30.03.3020