NFDI4BioDiversity: Biodiversity, Ecology & Environmental Data

Speaker: Frank Oliver Glöckner, MARUM - Center for Marine Environmental Sciences, University Bremen, contact@nfdi4biodiversity.org

Summary:

Biodiversity is more than just the diversity of living species. It includes genetic and phenotypic diversity of organisms, functional diversity, interactions and the diversity of populations and whole ecosystems. Mankind continues to dramatically impact the earth's ecosystem which is the foundation of human well-being. A general understanding of the status, trends, and drivers of biodiversity on earth is urgently needed to determine management options and devise conservation responses. Answers to scientifically and societal relevant questions can be found only through the availability of data integrated from multiple sources and building on different knowledge domains in science and society. This needs to be realized following the FAIR (Findable, Accessible, Interoperable, Re-usable) data principles. To foster easy access to interoperable data, NFDI4BioDiversity focuses on the following objectives:

Promoting research data management as an integral part of biodiversity research:

In an ideal world research data management (RDM) is seamlessly embedded within the complete scientific workflow. The insufficient co-design between biodiversity research and digitalisation has contributed to the fact that large parts of the current data management landscape are fragmented, ephemeral, not efficient, and lack quality. NFDI4BioDiversity will ensure that RDM becomes an integral, funded and credited part of biodiversity research, which is a prerequisite for successful data science approaches and a major requirement for the NFDI in general.

Enabling FAIRness of data:

As professional data management is still almost absent or underdeveloped in biodiversity science, it is often done according to individual, not necessarily FAIR compliant, workflows. Data for biodiversity research range from biolab experimental studies to manual and automatic device recording in the field. The data generation is often done by students and young researchers without much data management experience. The bandwidth of scientific questions and approaches in biodiversity and ecological communities aggravates the situation, also with respect to data sources and formats. As a speciality in biodiversity research a considerable amount of data is collected by citizen scientists and (semi-)professionals, which may not even have their data digitized. Reflected in our use cases and further addressed in our task areas NFDI4BioDiversity will support structured digital data acquisition, remove technical hurdles in data mobilisation, eliminate deficiencies, inconsistencies, and incompatibilities in data structures and semantics which impede effective interoperability and re-use of data. It will support education and training events to raise awareness and anchor the requirements in the biodiversity community. NFDI4BioDiversity will contribute to develop common minimum requirements standards across NFDI.

Consolidating FAIRness with quality:

Data of ambiguous quality will constrain the trust and acceptance of FAIRness. Since "quality of data" strongly depends on the application, the provenance of data must be conserved in the archiving and publication procedure. Data curation ensuring that data transformations are adequately documented in the metadata is therefore key for any re-use of data. Explicit definition of terms, formats and the use of controlled vocabularies and ontologies as well as clear licensing information must be ensured. We have set up measures to evaluate and prioritise repositories where quality assurance ranging from data provenance to data integrity

for long-term archival is implemented. Certification will be an important measure to gain trust in the community with respect to data deposition and re-use.

Embedding NFDI4BioDiversity into the national & international landscape:

As biodiversity science is not limited to national boundaries, integration of any future services into the existing international service landscape is crucial. Data and information must be exchanged independently of local, national or international borders. NFDI4BioDiversity will team up to coordinate international implementations and harmonise access to existing services. Structuring and leading the biodiversity community in Germany will boost the standing of Germany in generic initiatives like the European Open Science Cloud (EOSC).

NFDI-wide cross-cutting topics:

NFDI4BioDiversity has initiated an NFDI-wide discussion about cross-cutting topics. As a first result 21 NFDI consortia consolidated and agreed on a set of topics as outlined in the "Berlin Declaration on NFDI Cross-Cutting Topics (https://doi.org/10.5281/zenodo.3457213)". Over the last months the Berlin Declaration has been reshaped and widened to include all NFDI recommended and future consortia. The "Leipzig – Berlin Erklärung zu NFDI – Querschnittsthemen der Infrastrukturentwicklung" provides the new blueprint for our collaborative work in NFDI (https://doi.org/10.5281/zenodo.3895209).

NFDI4BioDiversity is particularly interested in contributing to the following cross-cutting topics:

1. **(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. We will follow the schema.org principles by building on a common core schema with community extensions like bioschemas.org.

2. **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 be solved only cooperatively, will be the mapping between terminologies and ontologies from different domains.

3. **Graduate Education**

NFDI4Health, NFDI4BioDiversity, NFDI4Earth together with the Federal State of Bremen and the University Bremen Research Alliance (UBRA), established a cross-domain graduate education programme on research data management and data science at the University of Bremen. This programme can 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.

4. **Governance & sustainability**

Suitable governance structures are key to ensure sustainable operations of a distributed infrastructure like NFDI. 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.

5. 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.

6. Research data commons

NFDI4BioDiversity is planning the implementation of the NFDI Research Data Commons (NFDI-RDC) to pool data and applications to leverage data science for the different stakeholders (Figure 1). 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.

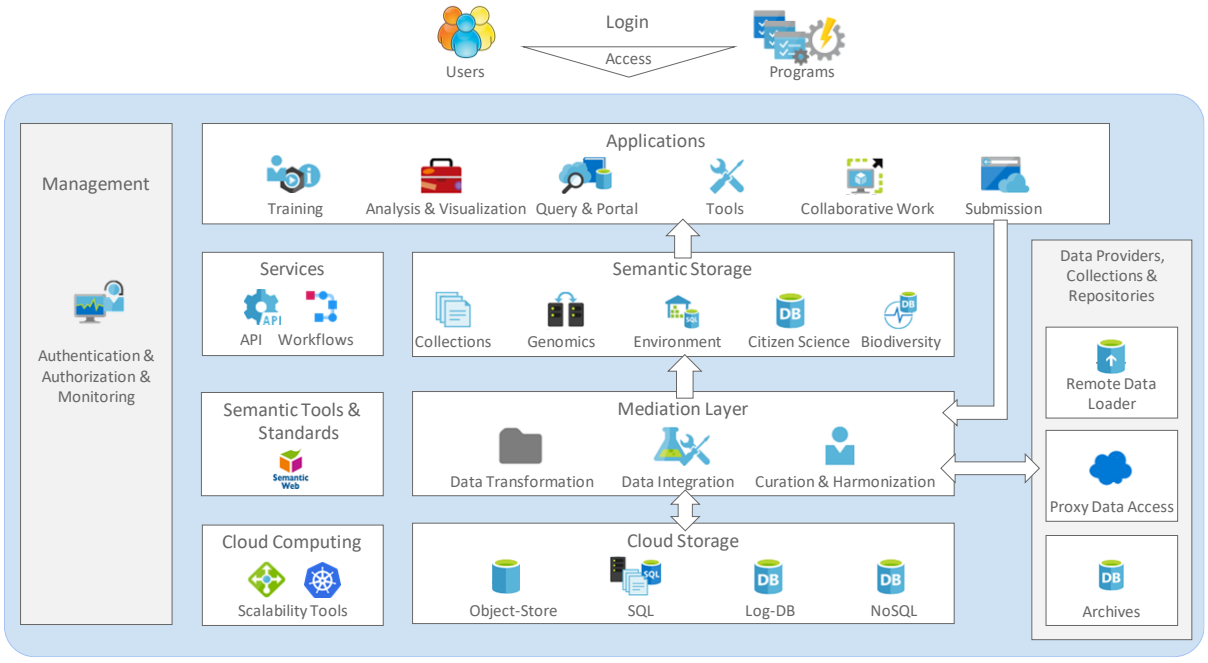

Figure1: Overall architecture of the NFDI-RDC.

Expectations for the NFDI Conference:

- Information exchange between recommended consortia and consortia applying in the 2nd and 3rd round
- Common recruitment activities
- Common infrastructures e.g. for compute and storage?
- Starting the implementation of cross-cutting topics, esp. metadata harmonisation and NFDI-RDC

Mitglieder des Konsortiums (Co-Sprecherinnen/Co-Sprecher und die weiteren, beteiligten Institutionen):

Co-spokespersons	Institution, location
Dr. Michael Diepenbroek	UniBremen-MARUM, Bremen
Dipl. Inf. Anton Güntsch	BGBM, Berlin
Prof. Dr. Jens Stoye	BIBI, Universität Bielefeld
Prof. Dr. Jörg Overmann	DSMZ, Braunschweig
Prof. Dr. Klaus Wimmers	FBN, Dummerstorf
Dr. Ivaylo Kostadinov	GFBio e.V., Bremen
Prof. Dr. Ramin Yahyapour	GWDG, Göttingen
Dr. Wolfgang Müller	HITS, Heidelberg
Dr. Uwe Scholz	IPK, Gatersleben
Dr. Dagmar Triebel	SNSB, München
Dr. Mark Frenzel	UFZ, Leipzig
apl. Prof. Dr. Birgit Gemeinholzer	UniGießen, Giessen
Prof. Dr. Alexander Goesmann	UniGießen, Giessen
Prof. Dr. Birgitta König-Ries	UniJena/iDiv, Jena
Prof. Dr. Aletta Bonn	UniJena/iDiv, Jena
Prof. Dr. Bernhard Seeger	UniMarburg, Marburg

Participants	Institution, location
Dr. Hubert Höfer	Arachnologischen Gesellschaft e.V. (AraGes-Spinnen), Putbus
Prof. Dr.Stephan Frickenhaus	Alfred Wegener Institute - Helmholtz Center for Polar- and Marine Research Bremerhaven (AWI), Bremerhaven
Dr. Janine Felden	Alfred Wegener Institute - Helmholtz Center for Polar- and Marine Research Bremerhaven (AWI), Bremerhaven
Dr. Jonas Zimmermann	Freie Universität Berlin, Botanic Garden and Botanical Museum Berlin, Research Group Diatoms (BGBM-DI), Berlin
Dr. Gerwin Kasperek	FID Biodiversitätsforschung an der Goethe Universität Frankfurt (BIOfid), Frankfurt
Dr. Gregor von Halem	Bundesamt für Seeschiffahrt und Hydrographie (BSH), Hamburg
Silvia Bender	"Bund für Umwelt und Naturschutz Deutschland e.V. – Friends of the Earth Germany (BUND), Berlin
Prof. Dr. Florian Leese	Aquatische Ökosystemforschung (COST Action), Universität Duisburg-Essen (DNAquaNet), Essen
Dr. Christoph Sudfeldt	Dachverband Deutscher Avifaunisten e.V. (DDA-Vögel), Münster
Klaus-Jürgen Conze	Gesellschaft deutschsprachiger Odonatologen e.V. (GdO-Libellen), Essen
Prof. Dr. Heiko Brunken	Gesellschaft für Ichthyologie e.V. (GfI-Fische), Solingen
Prof. Dr. Volkmar Wolters	Gesellschaft für Ökologie e.V. (GfÖ), Berlin
Prof. Dr. Helmut Hillebrand	Helmholtz Institute for Functional Marine Biodiversity at the University of Oldenburg (HIFMB), Oldenburg
Stephan Karger	Hessisches Landesamt für Natur, Umwelt und Geologie (HLNUG), Wiesbaden
Dr. Jana Moser	Leibniz Institut für Länderkunde (IfL), Leipzig

Participants	Institution, location
Prof. Dr. Hans-Peter Grossert	Leibniz-Institut für Gewässerökologie und Binnenfischerei (IGB), Berlin
Prof. Dr. Adrian Paschke	Institut für Angewandte Informatik (InfAI), Leipzig
Dr. Gotthard Meinel	Leibniz-Institut für ökologische Raumentwicklung (IÖR), Dresden
Dr. Holger Beer	Julius-Kühn-Institut Bundesforschungsinstitut für Kulturpflanzen (JKI), Quedlinburg
Dr. Kai Gedeon	Landesamt für Umweltschutz Sachsen-Anhalt (LAU),
Holger Lueg	Sächsisches Landesamt für Umwelt, Landwirtschaft und Geologie (LfULG), Dresden
Dr. Jana Hoffmann	Museum für Naturkunde Berlin (MfN), Berlin
Dr. Miguel D. Mahecha	Max-Planck-Institut für Biogeochemie (MPI-Biogeochem), Jena
Stefan Munzinger	Naturgucker gemeinnütziger e.G. (Naturgucker), Northeim
Dr. Steffen Caspari	Netzwerk Phytodiversität Deutschland (NetPhyD-Pflanzen), Schiffweiler
Prof. Dr. Jörg Müller	Nationalparkverwaltung Bayerischer Wald (Behörde im Geschäftsbereich des Bayerischen Staatsministeriums für Umwelt und Verbraucherschutz) (NP Bayerischer Wald), Grafenau
Jan Rommelfanger	Nationalparkamt Hunsrück-Hochwald (NP Hunsrück-Hochwald), Birkenfeld
Dr. Margit Ksoll-Marcon	Staatliche Archive Bayerns (SAB), München
Prof. Dr. Thomas Hickler	Senckenberg Gesellschaft für Naturforschung (SGN), Frankfurt am Main
Prof. Dr. Johanna Eder	Staatliches Museum für Naturkunde Stuttgart (SMNS), Stuttgart
Prof. Dr. Wolfram Horstmann	Niedersächsische Staats- und Universitätsbibliothek Göttingen (SUB), Göttingen
Florian Hoedt	Thünen Institute (Thünen-GDI), Braunschweig
Prof. Dr. Thomas Kneib	Universität Göttingen (UniGöttingen), Göttingen
Prof. Dr. Christian Wirth	Universität Leipzig (UniLeipzig), Leipzig
Dr. Peter Grobe	Zoologisches Forschungsmuseum Alexander Koenig (ZFMK), Bonn
Dr. Nicolas Dittert	Leibniz Zentrum für Marine Tropenforschung (ZMT), Bremen