

Bausteine Forschungsdatenmanagement
Empfehlungen und Erfahrungsberichte für die Praxis von
Forschungsdatenmanagerinnen und -managern

**NFDI4BioDiversity – NFDI-Konsortium für Biodiversitäts-,
Ökologie- und Umweltdatenⁱ**

Judith Sophie Weber

Barbara Ebert

Michael Diepenbroek

Ivaylo Kostadinov

Frank Oliver Glöckner

2021

Zitiervorschlag

Weber, Judith Sophie et al. 2021. NFDI4BioDiversity – NFDI-Konsortium für Biodiversitäts-, Ökologie- und Umweltdaten. *Bausteine Forschungsdatenmanagement. Empfehlungen und Erfahrungsberichte für die Praxis von Forschungsdatenmanagerinnen und -managern* Nr. 2/2021: S. 98-109. DOI: [10.17192/bfdm.2021.2.8334](https://doi.org/10.17192/bfdm.2021.2.8334).

Dieser Beitrag steht unter einer
[Creative Commons Namensnennung 4.0 International Lizenz \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/).

ⁱJudith Sophie Weber (ORCID: [0000-0001-8665-6382](https://orcid.org/0000-0001-8665-6382)), Barbara Ebert (ORCID: [0000-0003-3328-6693](https://orcid.org/0000-0003-3328-6693)), Michael Diepenbroek (ORCID: [0000-0003-3096-6829](https://orcid.org/0000-0003-3096-6829)), Ivaylo Kostadinov (ORCID: [0000-0003-4476-6764](https://orcid.org/0000-0003-4476-6764)), Frank Oliver Glöckner (ORCID: [0000-0001-8528-9023](https://orcid.org/0000-0001-8528-9023)).

1 Zusammenfassung

NFDI4BioDiversity ist ein Konsortium innerhalb der Nationalen Forschungsdateninfrastruktur (NFDI) mit einem Schwerpunkt auf Services für die Biodiversitätsforschung und Ökologie. Unzählige Studien belegen den Rückgang der Biodiversität auf unserem Planeten. Drängende Fragen in diesem Zusammenhang sind unter anderem, welche Auswirkungen dieser Rückgang der Biodiversität auf die betroffenen Ökosysteme hat und welche (anthropogenen) Einflüsse die Biodiversität reduzieren oder fördern. Um die komplexen Zusammenhänge zu verstehen und daraus Handlungsempfehlungen abzuleiten, werden vielfältige, qualitätsgesicherte und FAIR¹ Daten benötigt. Diese umfassen das Vorkommen von Pflanzen, Tieren und Mikroorganismen, einschließlich ihrer genotypischen, phänotypischen und funktionalen Vielfalt, sowie deren Zusammenspiel in Populationen und Ökosystemen. In NFDI4BioDiversity haben sich 49 Partnerorganisationen aus Wissenschaft, Behörden und Bürgerwissenschaften zusammengeschlossen, um die Mobilisierung und Publikation vorhandener Daten voranzutreiben, effiziente Workflows zu entwickeln und Tools und Services zur Unterstützung von Analysen bereitzustellen.

2 Einleitung

Biodiversität – oder biologische Vielfalt – hat als wissenschaftliches und umweltpolitisches Handlungsfeld national wie international große Bedeutung. Ob im Wasser, an Land oder in der Luft – je weniger Artenvielfalt es gibt, desto ärmer wird unser Planet und damit unser Leben. So vielfältig wie die Biodiversität selbst sind auch die Ansätze zu ihrer Erfassung, und ebenso heterogen sind die Datensätze, die daraus entstehen. Forschungsgruppen an Universitäten und Instituten beschäftigen sich mit dem Status der Biodiversität in unterschiedlichsten Ökosystemen sowie den Parametern und Größen, die diese beeinflussen: im Feld, im Labor oder anhand von Modellierungen. Hinzu kommen über Jahrhunderte systematisch gesammelte Daten und Objekte. Die vielfältigen Sammlungen der Naturkundemuseen bilden beispielsweise eine große Bandbreite der Artenvielfalt des Planeten ab und sind eine wichtige Grundlage der biologischen Systematik. Ähnlich bedeutsam ist der Bereich des staatlichen Naturschutzes mit seinen strukturierten Beobachtungsdaten. Einen besonderen Beitrag zur Biodiversitätsforschung leisten zudem die Bürgerwissenschaften, häufig in Verbänden oder hochspezialisierten Fachgesellschaften, die aus ihren unzähligen Beobachtungen wertvolle Datenbestände aufbauen. Ein Paradebeispiel ist die 2017 veröffentlichte Studie des Entomologischen Verein Krefeld e. V., die gesellschaftlich breite Aufmerksamkeit fand, da sie anhand von Zeitreihen erstmals das Ausmaß des so-

¹FAIR: Findable, Accessible, Interoperable, Reusable, aus: Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. et al. „The FAIR Guiding Principles for scientific data management and stewardship“, *Scientific Data* 3:160018 (March 2016), <https://doi.org/10.1038/sdata.2016.18>.

nannten Insektensterbens in Deutschland umfassend belegte².

NFDI4BioDiversity verfolgt das Ziel, zumindest einen Ausschnitt dieser heterogenen Datensätze und Datenquellen zu mobilisieren, zu strukturieren und zu standardisieren³. Das Konsortium kann dabei auf umfangreiche Vorarbeiten aufbauen. 2011 hatte eine Umfrage der DFG-Senatskommission für Biodiversitätsforschung ergeben, dass nur ein Bruchteil der in Forschungsprojekten erhobenen Daten für eine Nachnutzung überhaupt zugänglich war⁴. In der Folge wurde eine spezielle Förderung für die Erschließung wissenschaftlicher Daten beschlossen und es bildete sich die „German Federation for Biological Data – GFBio“⁵. Die Gruppe aus 20 institutionellen Partnern entwickelte über einen Zeitraum von sieben Jahren erste übergreifende Datendienste. Im Rahmen dieser Zusammenarbeit wurden die Datenzentren von Naturkundemuseen sowie etablierter Datenarchive wie PANGAEA und ENA an ein gemeinsames Portal angebunden⁶, Workflows für die Dateneinreichung vereinheitlicht und gemeinsame (Meta-)Datenstandards sowie Tools und Services rund um das Forschungsdatenmanagement (FDM) etabliert. NFDI4BioDiversity erweitert diesen Ansatz durch die Integration von Use Cases aus der Biodiversitäts-Community sowie Behörden und Bürgerwissenschaften.

3 Struktur von NFDI4BioDiversity

In NFDI4BioDiversity arbeiten 49 Partnerorganisationen aus ganz Deutschland zusammen. Repräsentiert ist der akademische Bereich mit Instituten und Universitäten, erweitert um Behörden, Fachgesellschaften und Verbände. NFDI4BioDiversity verfolgt dabei einen community basierten, von der Gemeinschaft gelenkten Ansatz. In fünf Aufgabenbereichen (Task Areas, kurz TA) wird auf die Ziele des Konsortiums hingearbeitet, wobei die TA5 sich mit der Koordination und der strategischen Ausrichtung des Gesamtkonsortiums befasst (**Abbildung 1**).

²Hallmann CA, Sorg M, Jongejans E, Siepel H, Hofland N, Schwan H, *et al.* (2017) More than 75 percent decline over 27 years in total flying insect biomass in protected areas. PLoS ONE 12(10): e0185809. <https://doi.org/10.1371/journal.pone.0185809>.

³Glöckner, F. O., Diepenbroek, M., Felden, J., Güntsch, A., Stoye, J., Overmann, J., Wimmers, K., Kostadinov, I., Yahyapour, R., Müller, W. Scholz, U., Triebel, D., Frenzel, M., Gemeinholzer, B., Goesmann, A., König-Ries, B., Bonn, A., Seeger, B. (2020, July 14). NFDI4BioDiversity - A Consortium for the National Research Data Infrastructure (NFDI). Zenodo. <http://doi.org/10.5281/zenodo.3943645>.

⁴DFG Strategiepapier, Zur Sicherung von umweltbezogenen biologischen Daten in Deutschland: http://www.dfg.de/download/pdf/dfg_im_profil/gremien/senat/biodiversitaet/ag_biodiversitatsdaten_datensicherung.pdf.

⁵Diepenbroek M, Glöckner F, Grobe P, Güntsch A, Huber R, König-Ries B, Kostadinov I, Nieschulze J, Seeger B, Tolksdorf R, & Triebel D (2014). Towards an Integrated Biodiversity and Ecological Research Data Management and Archiving Platform: The German Federation for the Curation of Biological Data (GFBio). Plödereder E, Grunske L, Schneider E & Ull D (eds): Informatik 2014 – Big Data Komplexität meistern. GI-Edition: Lecture Notes in Informatics (LNI) – Proceedings 232: 1711-1724. Köllen Verlag, Bonn. https://species-id.net/s/media/6/61/Gfbio_informatik2014.pdf.

⁶<https://www.gfbio.org/>, abgerufen am 29.01.2021.

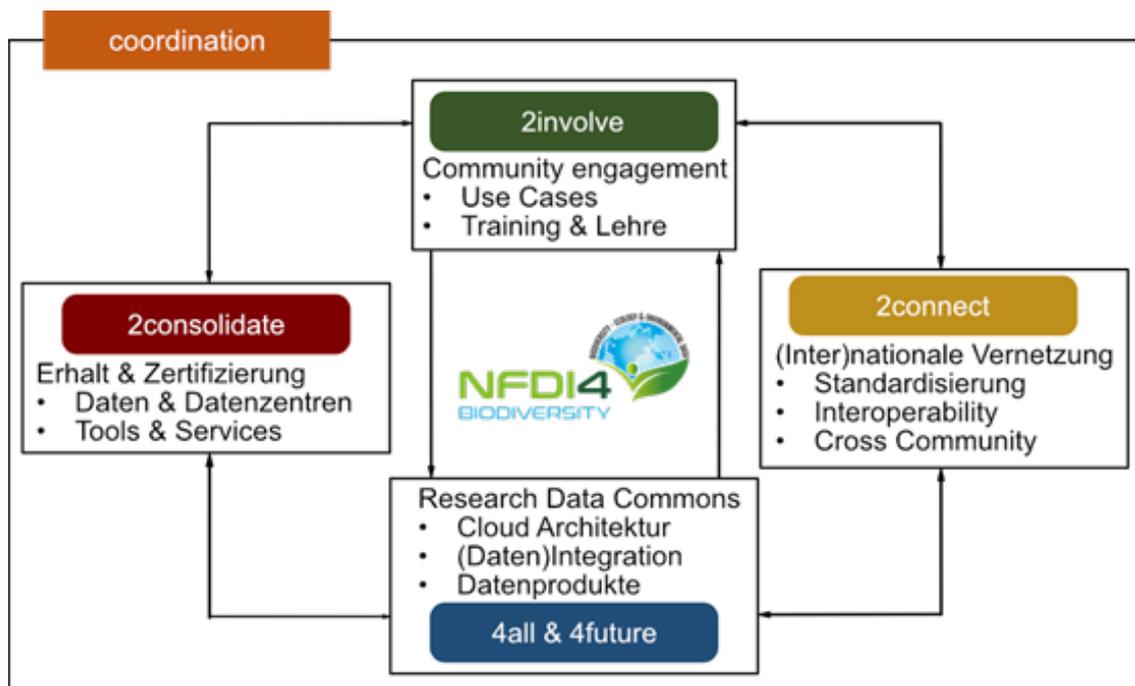


Abbildung 1: Die Struktur von NFDI4BioDiversity: Die 5 Arbeitspakete besitzen die Schwerpunkte Community engagement (*2involve*), nationale und internationale Vernetzung (*2connect*), langfristige Bewahrung und Zertifizierung von Tools und Daten (*2consolidate*) und Research Data Commons (*4all & 4future*), sowie die Gesamtkoordination. Durch einen engen Austausch zwischen den Arbeitsbereichen werden die Bedürfnisse und Expertisen der einzelnen Fokusthemen jeweils in den Arbeiten der anderen Bereiche berücksichtigt. Darstellung angepasst aus Glöckner *et al.*, (2019), vgl. FN 3.

TA 1 (*2involve*) steht für die breite Einbindung der Biodiversitäts-Community. In 23 Use Cases werden dort konkrete Anforderungen des Datenmanagements unterschiedlicher Umgebungen bearbeitet. Dies umfasst zum Beispiel die Entwicklung und Anpassung von Software-Lösungen für die Repräsentation der Daten bei den Partnern. Zudem ist in TA1 Lehre und Training verankert. Neben NutzerInnenschulungen für die bereitgestellten Tools und Services liegt ein starker Fokus auf dem Transfer von Materialien zur praktischen Anwendung in der wissenschaftlichen Ausbildung. So ist beispielsweise die Bereitstellung von Trainingsdatensätzen geplant, die in fachbezogenen Studiengängen zum Einsatz kommen können. Am Standort Bremen beteiligt sich NFDI4BioDiversity bereits an „Data Train“, einer Lehrreihe mit den Themen Datenmanagement und Data Science der U Bremen Research Alliance, die gemeinsam mit Akteuren aus vier in Bremen vertretenen NFDI-Konsortien entwickelt wurde⁷.

⁷NFDI4BioDiversity, NFDI4Health, KonsortSWD und NFDI4Ing, siehe <https://www.uni-bremen.de/research-alliance/forschungsdaten/data-train> (abgerufen am 29.01.2021).

TA2 (*2connect*) bündelt die Vernetzungsaktivitäten des Konsortiums in der nationalen und internationalen Landschaft und umfasst insbesondere Standardisierungsinitiativen, aber auch die konsortienübergreifenden sogenannten Querschnittsthemen und die Realisierung von Synergien innerhalb der NFDI.

Im Zentrum von TA3 (*2consolidate*) steht die Bewahrung und Zertifizierung von Daten, Datenzentren, Tools und Services, die durch die Partner von NFDI4BioDiversity entwickelt oder von extern angegliedert werden.

Die Entwicklung und Etablierung der Research Data Commons (RDC), einer cloudbasierten Infrastruktur, ist die Aufgabe der TA4 (*4all&4future*). Es ist geplant, die NFDI-RDC in Zusammenarbeit mit weiteren Konsortien zu entwickeln, als gemeinsame Infrastruktur für den Zugriff, die Suche, das Zusammenführen und die Analyse heterogener Daten aus verschiedenen Quellen.

4 Welche Daten sind relevant für NFDI4BioDiversity?

Im Kern von NFDI4BioDiversity stehen Daten über die Artenvielfalt der Tiere, Pflanzen und Mikroorganismen. Diese werden auf verschiedenste Weise erfasst, zum Beispiel als Beobachtungsdaten, die in Form von Tabellen, Fotos, Videos oder Audiodateien festgehalten werden. Diese Beobachtungen können einmalig in Raum und Zeit stattfinden (z. B. während eines Spaziergangs oder einer Feldexkursion) oder Teil gezielter Langzeitstudien sein. Häufig werden auch Exemplare gesammelt und zur weiteren Analyse bzw. Beschreibung konserviert. Hinzu kommt eine wachsende Menge an genetischer Information zu den beobachteten Spezies, also Sequenzdaten – vom sogenannten Barcoding (einer Methode zur Artenbestimmung über genetische Marker) über Genomsequenzen bis hin zu Metagenomanalysen. Ebenso von Interesse sind die Daten anderer -omics Disziplinen wie z. B. Metabolomics, Glycomics und Transcriptomics, die Informationen über funktionelle Zusammenhänge bieten.

Daneben sind Daten zu den Umweltbedingungen im Lebensraum der Arten, wie auch über die Landnutzung und die Besiedlung relevant. Durch den verstärkten Einsatz von Sensoren, die bestimmte Umweltparameter über die Zeit messen und laufend Daten liefern (Streaming), wächst die Menge erfasster und zu nutzender Daten ständig⁸.

Diese Daten sind nicht nur für die Community der NFDI4BioDiversity von Interesse, sondern liefern auch in anderen thematischen Zusammenhängen wertvolle Informationen. Ebenso werden für die Biodiversitätsforschung essentielle Daten innerhalb anderer Disziplinen erhoben. Im Wissenschaftsbereich finden sich zum Beispiel breite Überschneidungen mit den Erd- und Umweltwissenschaften, der Agrarwissenschaft, und der Kulturgeographie. Gesamtgesellschaftlich besteht traditionell eine enge Verzahnung zwischen der Biodiversitätsforschung und dem Umwelt- und Naturschutz (Biodiversitätsmonitoring). Daher besteht innerhalb des Konsortiums und darüber hinaus hohes Interesse an einer standardisierten Strukturierung und Beschreibung der Daten, die ihre spätere (Nach-)Nutzbarkeit optimiert. Besonders relevant ist hier die

⁸Ein Beispiel ist das AMMOD Projekt: <https://ammod.de/>, abgerufen am 29.01.2021.

internationale Vernetzung, zum Beispiel mit der europäischen Forschungsinfrastruktur für die Lebenswissenschaften (ELIXIR⁹, über den Knoten des Deutschen Netzwerks für Bioinformatik Infrastruktur de.NBI¹⁰) und der Global Biodiversity Information Facility GBIF¹¹, um nur zwei der zahlreichen kooperierenden Netzwerke zu nennen.

5 Mobilisierung, Archivierung und Publikation der Daten

Die bereits in GFBio angebotenen Datenzentren der Naturwissenschaftlichen Sammlungen (Tabelle 1), sowie die drei auf digitale Daten spezialisierten Forschungsdatenrepositorien (Tabelle 2) spielen innerhalb von NFDI4BioDiversity eine zentrale Rolle als Datenanbieter und Dienstleister. Im Rahmen der strukturierten Langzeitarchivierung und Datenpublikation leisten sie intensive Kuration und Standardisierung der Datensätze durch Fachleute. Dies sorgt für fassbare Mehrwerte im Hinblick auf eine spätere wissenschaftliche Nachnutzung und ermöglicht beispielsweise die Kombination von Datensätzen verschiedener Provenienz.

Im Laufe des Aufbaus von NFDI4BioDiversity soll der Kreis der Datenzentren sukzessive erweitert werden. Innerhalb der Use Cases werden außerdem wertvolle Datenbestände aus den weiteren Partnerorganisationen erschlossen („mobilisiert“), sodass sie Teil der geplanten Research Data Commons (RDC) werden können, einer Cloud-Infrastruktur, in deren Rahmen sie für breitere Anwendungen zur Verfügung stehen. Die technische Anbindung dieser Ressourcen wird eine der Aufgaben während der Projektlaufzeit sein. Die perspektivische Anbindung von weiteren Datenquellen wird besonders in Zusammenarbeit mit anderen NFDI-Konsortien angestrebt, um die Vereinheitlichung und Vereinfachung von Workflows noch weiter in die Communitys hineinzutragen, und die größtmögliche Menge relevanter Daten für die NFDI Research Data Commons zu mobilisieren.

Standardisierung und Qualitätssicherung werden dabei nicht auf die Datenebene beschränkt bleiben, sondern betreffen auch die Prozesse. Im Rahmen von GFBio wurde der Zertifizierungsprozess zum Erwerb des Core Trust Seal¹² für vertrauenswürdige Repositorien bereits für die sieben bestehenden Datenzentren der naturwissenschaftlichen Sammlungen (Tabelle 1) und das im Aufbau begriffene Jena Data Center in die Wege geleitet. Das Ziel von NFDI4BioDiversity ist es, weitere Betreibereinrichtungen zur Zertifizierung zu begleiten, sofern sich ihre Dienste dafür qualifizieren.

⁹<https://elixir-europe.org>, abgerufen am 29.01.2021.

¹⁰<https://www.denbi.de>, abgerufen am 29.01.2021.

¹¹<https://www.gbif.org>, abgerufen am 29.01.2021.

¹²<https://www.coretrustseal.org>, abgerufen am 29.01.2021.

Tabelle 1: Überblick über die im Rahmen von GFBio bereits angeschlossenen Datenzentren der Naturwissenschaftlichen Sammlungen. Quelle: eigene Darstellung, gestützt durch die mitwirkenden Datenzentren.

Abkürzung	Name	Fokus	Logo
BGBM	Freie Universität Berlin, ZE Botanischer Garten und Botanisches Museum Berlin ^{13,14}	Botanik	
DSMZ	Leibniz-Institut DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen ^{15,16}	Mikroorganismen und Zelllinien	
MfN	Museum für Naturkunde – Leibniz- Institut für Evolutions- und Biodiversitätsforschung ^{17,18}	Zoologie, Ökologie, Paläontologie	
SGN	Senckenberg Gesellschaft für Naturforschung ^{19,20}	Botanik, Zoologie, Paläontologie, Anthropologie	
SMNS	Naturkundemuseum Stuttgart ^{21,22}	Botanik, Zoologie, Paläontologie	
SNSB	Staatliche Naturwissenschaftliche Sammlungen Bayerns ^{23,24}	Botanik, Mykologie, Zoologie, Paläontologie	
ZFMK	Zoologisches Forschungsmuseum Alexander Koenig – Leibniz- Institut für Biodiversität der Tiere ^{25,26}	Zoologie	

¹³<https://www.bgbm.org/de>, abgerufen am 11.05.2021.

¹⁴<https://www.gfbio.org/data-centers/bgbm>, abgerufen am 11.05.2021.

¹⁵<https://www.dsmz.de>, abgerufen am 11.05.2021.

¹⁶<https://www.gfbio.org/data-centers/dsmz>, abgerufen am 11.05.2021.

¹⁷<https://www.museumfuernaturkunde.berlin/de>, abgerufen am 11.05.2021.

¹⁸<https://www.gfbio.org/data-centers/mfn>, abgerufen am 11.05.2021.

¹⁹<https://www.senckenberg.de>, abgerufen am 11.05.2021.

²⁰<https://www.gfbio.org/data-centers/sgn>, abgerufen am 11.05.2021.

²¹<https://naturkundemuseum-bw.de>, abgerufen am 11.05.2021.

²²<https://www.gfbio.org/data-centers/smns>, abgerufen am 11.05.2021.

²³<https://www.snsb.de/index.php/de>, abgerufen am 11.05.2021.

²⁴<https://www.gfbio.org/data-centers/snsb>, abgerufen am 11.05.2021.

²⁵<https://www.zfmk.de>, abgerufen am 11.05.2021.

²⁶<https://www.gfbio.org/data-centers/zfmk>, abgerufen am 11.05.2021.

Tabelle 2: Überblick über die im Rahmen von GFBio bereits angebotenen Forschungsdatenrepositorien (nur digitale Objekte). Quelle: eigene Darstellung, gestützt durch die mitwirkenden Datenzentren.

Abkürzung	Name	Fokus	Logo
ENA	European Nucleotide Archive ^{27,28}	Nucleotide Sequenzdaten	
e!DAL-PGP	Plant Genomics & Phenomics Research Data Repository ^{29,30}	Phänotypische und genomische Pflanzendaten	
PANGAEA	Data Publisher for Earth & Environmental Science ^{31,32}	Geologie, Umweltdaten	

6 Der Status Quo

NFDI4BioDiversity baut auf den bestehenden Services von GFBio auf. Diese umfassen für DatenerzeugerInnen zum Beispiel die individuelle Beratung in Fragen des FDM und der Erstellung eines Datenmanagement-Plans (DMP)³³ sowie ein einheitliches Data Submission System³⁴, über das Sequenzdaten und andere Outputs aus Forschungsprojekten zentral eingereicht und über GFBio an die jeweils geeigneten Datenzentren vermittelt werden.

Für DatennachfragerInnen gibt es ein Datenportal³⁵, in dem quer durch alle angeschlossenen Datenzentren recherchiert werden kann, sowie ein „Visualisation, Analysis & Transformation“ (VAT) System³⁶, mit dem die recherchierten Daten räumlich-zeitlich aufbereitet werden können. Das Service-Portfolio wird ergänzt durch einen Terminologie-Service³⁷ und die enge Zusammenarbeit mit zwei Software-Tools für das laufende Datenmanagement im Forschungsprozess (Diversity Workbench³⁸ und BEXIS³⁹). Jede DMP-Beratung, Dateneinreichung oder sonstige Anfrage läuft über einen zentralen Help Desk und generiert ein Ticket, das von einem eingearbeiteten Team bearbeitet wird. Repräsentanten der Datenzentren sind Teil des Help Desks, so dass

²⁷<https://www.ebi.ac.uk/ena/browser/home>, abgerufen am 11.05.2021.

²⁸<https://www.gfbio.org/data-centers/ena>, abgerufen am 11.05.2021.

²⁹<https://edal-pgp.ipk-gatersleben.de>, abgerufen am 11.05.2021.

³⁰<https://www.gfbio.org/data-centers/e-dal-pgp>, abgerufen am 11.05.2021.

³¹<https://www.pangaea.de>, abgerufen am 11.05.2021.

³²<https://www.gfbio.org/data-centers/pangaea>, abgerufen am 11.05.2021.

³³<https://www.gfbio.org/dmpt>, abgerufen am 29.01.2021.

³⁴<https://submissions.gfbio.org>, abgerufen am 29.01.2021.

³⁵<https://www.gfbio.org/search>, abgerufen am 29.01.2021.

³⁶<https://vat.gfbio.org>, abgerufen am 29.01.2021.

³⁷<https://terminologies.gfbio.org>, abgerufen am 29.01.2021.

³⁸https://diversityworkbench.net/Portal/Diversity_Workbench, abgerufen am 29.01.2021.

³⁹<https://fusion.cs.uni-jena.de/bpp>, abgerufen am 29.01.2021.

fachspezifische Fragen und Dateneinreichungen durch ExpertInnen bearbeitet werden können⁴⁰.

7 Die Use Cases als Kern von NFDI4BioDiversity

Die Daten aus der hier beschriebenen, sehr heterogenen Akteurslandschaft können nicht auf einen Schlag integriert werden. Im Konsortium wurden daher Use Cases definiert, die Ausschnitte dieser Landschaft repräsentieren bzw. die Anbindung wichtiger, noch ausstehender nationaler und internationaler Gruppen umfassen.

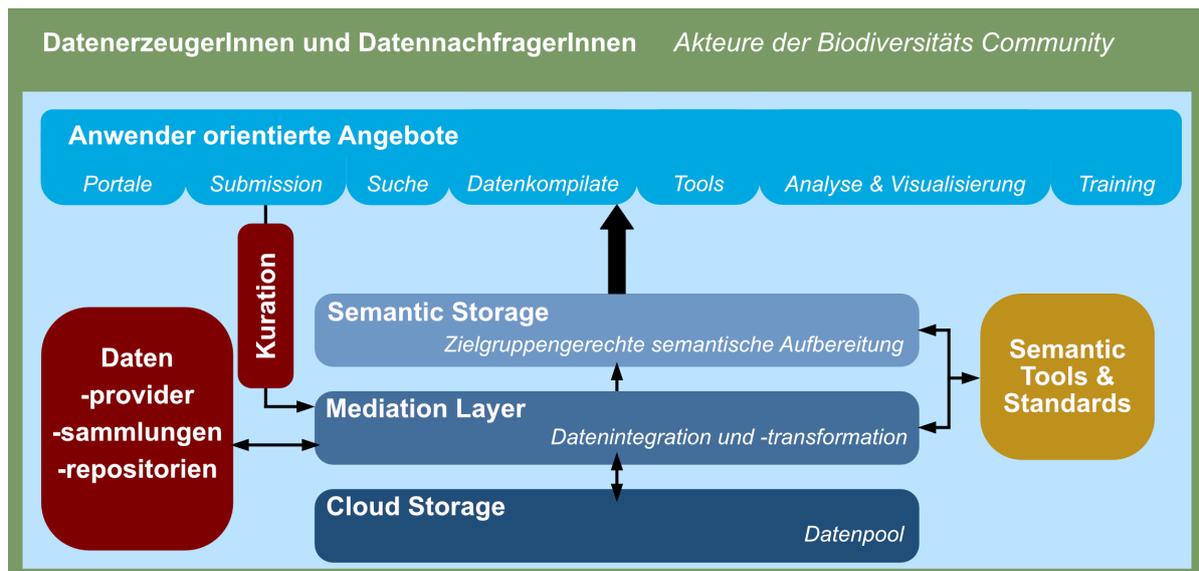


Abbildung 2: Schematischer Aufbau der Research Data Commons. Eigene Darstellung, angelehnt an Glöckner *et al.*, (2019), vgl. FN 3.

Dazu gehören beispielsweise die europäische eLTER Forschungsinfrastruktur⁴¹ mit ihren ökologischen Langzeitdaten, einige auf spezielle Tier- bzw. Pflanzenfamilien spezialisierte Fachgesellschaften sowie Akteure im Bereich des Umwelt- und Naturschutzes und der Bürgerwissenschaften.

Während die Visionen der „Open Science“ in der Politik und Gesellschaft von frei fließenden, FAIRen Daten handeln, sind die Probleme auf der Arbeitsebene meist andere. In einigen Use Cases wird Unterstützung bei der semantischen Anreicherung von (Meta-)Daten benötigt, etwa durch Bereitstellung von Artenreferenzlisten oder eine strukturierte Ablage von Beobachtungsdaten. Andere Bedarfe betreffen grundlegende

⁴⁰Für eine detaillierte Beschreibung der GFBio Services siehe Astor, T., Weber, J., Kostadinov, I., Glöckner, F. O., Nieschulze, J. und das GFBio Team. 2021. Potential für ein starkes Netzwerk zwischen GFBio und FDM-Beratenden an Universitäten und Forschungsinstituten. *Bausteine Forschungsdatenmanagement*, Nr. 1 (März). German:22-31.<https://doi.org/10.17192/bfdm.2021.1.8311>

⁴¹<https://www.lter-europe.net/lter-europe>, abgerufen am 29.01.2021.

IT-Dienstleistungen, zum Beispiel die Entwicklung von Datenbanklösungen oder Web-Portalen, oder der Bereitstellung von bereits vorhandenen Daten über standardisierte Schnittstellen. Die Bearbeitung der Use Cases in der partnerschaftlichen Struktur des NFDI4BioDiversity-Konsortiums befördert dabei die Kooperation unter den beteiligten Parteien, aber auch die Verbindungen zu anderen NFDI-Konsortien, etwa hinsichtlich der Etablierung gemeinsamer Standards und Workflows.

8 Die Vision der NFDI Research Data Commons (RDC)

Wie bereits skizziert, sind die meisten der erhobenen (Forschungs-)Daten disziplinübergreifend bzw. über den Sektor Wissenschaft hinaus gesellschaftlich relevant. Je nach Fragestellung und Forschungsmethodik werden zielgruppengerechte Datenkompile und Datenprodukte benötigt⁴². Um möglichst große Mehrwerte aus den bereitgestellten Daten ziehen zu können, braucht es eine Basis-Infrastruktur, die es verschiedensten Akteuren erlaubt, a) auf eine große Menge heterogener Daten (unterschiedlicher Disziplinen) zuzugreifen (Mobilisierung), b) diese in eine vergleichbare Struktur zu bringen (Integration und Transformation), und sie c) anlassbezogen und zielgruppengerecht aufzubereiten (Harmonisierung und Provenienz). Diese Vision treibt die Entwicklung der Research Data Commons (RDC), einer Art wissenschaftlichen Allmende. Die NFDI Research Data Commons sind das technische Rahmenkonzept für eine Vielfalt von intermediären Diensten, die zwischen den Bedürfnissen der Datennachfrage und der Erzeugerseite vermitteln. Diese werden in verschiedene „Schichten“ geordnet, die der Erzeugerseite eine Ankopplung von Beständen und Endnutzenden einen flexiblen Zugriff auf relevante Daten verschiedener Quellen und zugehörige Dienstleistungen ermöglichen soll (Abbildung 2).

Im „Mediation Layer“ erfolgt die technische Anbindung der fachlich ausdifferenzierten Datenquellen und der dort verfügbaren Daten sowie die Transformation dieser heterogenen Daten, sodass diese integrierbar und durchsuchbar sind. Darüber ist eine semantische Ebene geplant, die eine semantische Anreicherung und Beschreibung der Daten ermöglicht, sodass Daten aus verschiedenen Disziplinen in eine gemeinsame Sprache übersetzt für verschiedene Zielgruppen aufbereitet werden können. In der Applikationsschicht (Anwendungen) finden Nutzende benutzerfreundliche Produkte wie Datenportale, Dashboards und andere Applikationen, sodass die harmonisierten Daten abgerufen, mit RDC-kompatiblen, standardisierten Tools analysiert oder auch mit eigenen Daten integriert werden können. Angestrebt wird eine Art Ökosystem an nutzergenerierten Applikationen. Auf dieser Ebene sind z. B. auch Dateneinreichungen möglich, um Forschungsdatensätze aus der Individual- und Verbundforschung bei einem der angebundenen Datenzentren zu archivieren und zu publizieren.

⁴²Baker, Karen S.; Duerr, Ruth E.; Parsons, Mark A. (2015): Scientific Knowledge Mobilization. Co-evolution of Data Products and Designated Communities. In: International Journal of Digital Curation 10 (2), S. 110–135. <https://doi.org/10.2218/ijdc.v10i2.346>.

Technisch werden die Research Data Commons als Multi-Cloud aufgesetzt, um eine hohe Flexibilität und Skalierbarkeit zu gewährleisten. Auch hybride Konstruktionen und Edge-Komponenten sind vorgesehen. Geplant ist die Nutzung bereits etablierter Cloud-Computing-Ressourcen aus dem Wissenschaftsbereich, zum Beispiel die des deutschen Netzwerks für Bioinformatik (de.NBI). Die einzelnen Komponenten werden modular aufgebaut und über standardisierte Schnittstellen (e.g. REST, gRPC) miteinander kommunizieren. Die Bereitstellung wird mithilfe von (Docker) Container erfolgen, die über Kubernetes verwaltet werden.

Die RDC-Entwicklung ist als NFDI-Querschnittsthema definiert.^{43,44} Die in der ersten Runde geförderten NFDI-Konsortien haben Interesse bekundet, sich in die Entwicklung einer gemeinsamen interoperablen Infrastruktur einzubringen⁴⁵, zudem liegen Interessensbekundungen weiterer Konsortialinitiativen vor, die im Frühjahr 2021 noch im Antragsprozess sind (NFDI4Earth, NFDI4Agri, NFDI4Microbiota).

9 Herausforderungen

Der Aufbau einer nationalen Forschungsdateninfrastruktur als ein „Netzwerk der Netzwerke“ ist für alle Beteiligte ein forderndes soziotechnisches Experiment. Speziell für NFDI4BioDiversity mit seinen breit aufgestellten Partnerorganisationen bedeutet dies einen Brückenschlag zwischen wissenschaftlichen und zivilgesellschaftlichen Akteuren, deren Interessen der Datennutzung überein gebracht werden müssen. Die Mobilisierung von Daten ist schon innerhalb der Wissenschaft kein leichtes Geschäft, wo viele Forschende und Forschungsgruppen verständlicherweise hohen Wert auf die Kontrolle über ihre Daten legen und der Aufwand für deren Erschließung und Publikation noch wenig gewürdigt wird.

Mit Blick auf die Landschaft der Datenpublikationsplattformen gilt es zudem, auf eine Differenzierung hinsichtlich der wissenschaftlichen Reputation hinzuwirken. Die Einreichung eines Datensatzes zur Aufnahme in fachlich kuratierte Datensammlungen wie die von NFDI4BioDiversity geht mit Selektion einher und ist aufgrund der Anforderungen an Qualität und Dokumentation ein anspruchsvoller wissenschaftlicher Prozess. Die Aufnahme eines Datensatzes in solch eine Sammlung muss sich als Leistung in den Anerkennungsmechanismen des Wissenschaftsbetriebs noch deutlicher abbilden.⁴⁶

⁴³Glöckner, F. O., Diepenbroek, M., Felden, J., Overmann, J., Bonn, A., Gemeinholzer, B., et al. (2019, September 22). Berlin Declaration on NFDI Cross-Cutting Topics (Version 1). Zenodo. <http://doi.org/10.5281/zenodo.3457213>.

⁴⁴Bierwirth, M., Glöckner, F. O., Grimm, C., Schimmler, S., Boehm, F., Busse, C. et al. (2020, June 15). Leipzig-Berlin-Erklärung zu NFDI-Querschnittsthemen der Infrastrukturentwicklung. Zenodo. <http://doi.org/10.5281/zenodo.3895209>.

⁴⁵Ebert, B., Fluck, J., Glöckner, F. O., Koepler, O., Miller, B., Schmitt, R., et al. (2021, March 11). NFDI Cross-cutting Topics Workshop Report. Zenodo. <http://doi.org/10.5281/zenodo.4593770>.

⁴⁶Vgl. Empfehlungen 4.2 und 4.4 des Rats für Informationsinfrastrukturen *In: RfII – Rat für Informationsinfrastrukturen: Herausforderung Datenqualität – Empfehlungen zur Zukunftsfähigkeit von Forschung im digitalen Wandel*, zweite Auflage, Göttingen 2019, 172 S.; <http://www.rfii.de/?p=4043>.

Bei den außerwissenschaftlichen Partnern gilt es, die Mobilisierung der vorhandenen Daten für die Forschung mit konkreten Mehrwerten für deren eigenes Tagesgeschäft zu verbinden. Das kann auch in der Beschleunigung einer ohnehin anstehenden Digitalisierung von Arbeitsprozessen liegen. Die Bemühungen um eine Konsolidierung des nationalen Biodiversitätsmonitorings, die mit der Gründung neuer Datenzentren und parallelen Vernetzungsinitiativen einhergeht, befördert unser Ziel der wirksamen Datenmobilisierung über die Grenzen der Sektoren hinaus⁴⁷. Solch kraftvolle Anstrengungen seitens der zuständigen Ressorts bei Bund und Ländern bedeuten, dass die NFDI4BioDiversity sich sowohl in der Wissenschafts- als auch der Umweltpolitik gut platzieren muss, um als Akteursnetzwerk Anerkennung zu finden. Erfolgsfaktoren werden die Organisation eines guten und vor allem konstanten transdisziplinären Austausches sein sowie die Bereitstellung von konkreten, nützlichen und ansprechenden Services und Tools, die eine Vielzahl von Anwendungen mit und in der Community unterstützen.

10 Ausblick

Das NFDI4BioDiversity-Konsortium startet auf einer guten Grundlage. Aus dem GFBio-Netzwerk heraus sind Services für den „Long Tail of Science“ etabliert, die Daten aus der Individual- und Verbundforschung erschließen helfen und Dienstleistungen für das Forschungsdatenmanagement von Arbeitsgruppen innerhalb der Biodiversitätsforschung anbieten. Dieser Ansatz wird in NFDI4BioDiversity signifikant erweitert. Durch Anbindung von amtlicherseits erhobenen Daten und den Bürgerwissenschaften werden Biodiversitätsdaten nicht nur aus dem akademischen Umfeld, sondern aus der Breite der Akteurslandschaft heraus mobilisiert. Im Gegenzug gilt es, fassbare Mehrwerte für Wissenschaft, Politik und Gesellschaft zu generieren. Mit anderen Worten: Wir freuen uns, gemeinsam mit dem NFDI-Direktorat durchzustarten und das Großvorhaben „Nationale Forschungsdateninfrastruktur“ tatkräftig zu befeuern.

Danksagung

Die Autorinnen und Autoren des Artikels haben zur Erstellung dieses Porträts verschiedene Vorarbeiten aus den GFBio und NFDI4BioDiversity Netzwerken genutzt. Referenzen wurden, soweit möglich, gesetzt. Darüber hinaus möchten wir besonders allen beteiligten Personen danken, die die hier dargestellten Strukturen und Ergebnisse in den vergangenen acht Jahren mitgestaltet und -getragen haben und in den nächsten Jahren daran mitarbeiten werden, die Pläne innerhalb von NFDI4BioDiversity zu verwirklichen.

⁴⁷Vgl. das kommentierte Grobkonzept für das nationale Monitoringzentrum zur Biodiversität <https://www.bmu.de/download/grobkonzept-fuer-das-nationale-monitoringzentrum-zur-biodiversitaet/>.