

NFDI4Life Umbrella Forschungsdateninfrastruktur für die Lebenswissenschaften

NFDI4Life Umbrella

Sprecher: Prof. Dr. Dietrich Rebholz-Schuhmann, ZB MED Informationszentrum
Lebenswissenschaften, rebholz-schuhmann@zbmed.de

Forschungsgebiet NFDI4Life Umbrella richtet sich an alle übergreifenden Erfordernisse des Forschungsdatenmanagements in den Lebenswissenschaften mit all seinen Fächern. Die derzeit vorgeschlagenen disziplinspezifischen Konsortien im Bereich der Lebenswissenschaften (NFDI4BioDiversity, NFDI4Health, NFDI4Med, NFDI-Neuro, NFDI4Microbiota, NFDI4AIRR, NFDI4Agri, NFDI4Chem und DataPlant) streben an, ihre Kräfte im Rahmen von NFDI4Life Umbrella zu bündeln und betonen den Bedarf für ein solches koordinatives Netzwerk aufgrund (1) der zunehmend interdisziplinären und translationalen Forschungsdaten und dem Informationsbedarf im Bereich der Lebenswissenschaften, (2) spezifischer, aber übergeordneter und bereichsübergreifender Aufgaben für die Lebenswissenschaften (z.B. übergreifende Standards, Methoden und rechtliche Rahmenbedingungen, Vernetzung in der Wissenschaftsgemeinschaft als Ganzes, kultureller Wandel, Reputationsbildung oder politische Interessensvertretung), die in einem lebenswissenschaftlichen Dachkonzept besser adressiert und entwickelt werden können als in den einzelnen Fachkonsortien (3) dem hohen Grad der Verknüpfung von lebenswissenschaftlichen Daten aufgrund ähnlicher wissenschaftlicher Prinzipien, d.h. dieselben Mechanismen und angewandten Methoden liegen hier zugrunde, die sowohl beim Menschen als auch bei Tieren oder Pflanzen, also in der biologischen, medizinischen, epidemiologischen, veterinärmedizinischen, landwirtschaftlichen, ökologischen und Biodiversitätsforschung relevant sein können, (4) ähnlicher Anforderungen an die Forschungsdateninfrastruktur für alle Organismen, die sich in gemeinsamen Datenstrukturierungsbedürfnissen widerspiegeln (z.B. für genomische Sequenzierung, Proteom- und Metabolomikdaten, Bildgebung).

Ziele NFDI4Life Umbrella definiert daher folgende Ziele (1) Sicherstellung von FAIRen Daten über die NFDI-Konsortien im Bereich der Lebenswissenschaften, (2) Standardisierung von bereichsübergreifend verwendeten Daten und Prozessen, insbesondere von relevanten Metadatenstrukturen, sowie Harmonisierung von Standards im gesamten Bereich der Lebenswissenschaften und Koordination der Prozesse, (3) gemeinsame Nutzung von Cloud-Infrastrukturen und IT-Lösungen, (4) konsortial übergreifende Ansprache von Fachcommunities im Bereich der Lebenswissenschaften durch konzertierte Aus- und Weiterbildung, internationale Vertretung und Sichtbarkeit der deutschen lebenswissenschaftlichen Datengemeinschaft, (5) Organ der lebenswissenschaftlichen NFDI-Konsortien gegenüber Politik, Förderern, der Allianz der Wissenschaftsorganisationen, anderen NFDI-Konsortien und internationalen Organisationen wie Go FAIR, RDA oder EOSC, (5) Verknüpfung der bestehenden Initiativen in Bezug auf die Verwaltung von Forschungsdaten aus Fachcommunities und Informationsinfrastrukturen.

Anforderungen an das Forschungsdatenmanagement Im Hinblick auf den aktuellen Stand des Forschungsdatenmanagements in den Lebenswissenschaften ergeben sich Bedarfe für ein koordinierendes Netzwerk sowohl aus Sicht der Forschung als auch der Infrastrukturbetreiber. Die Bedürfnisse der Nutzer konzentrieren sich auf Forschungsfragen,

die die Kombination von Datenquellen aus verschiedenen Bereichen der Lebenswissenschaften erfordern. NFDI4Life Umbrella wird solche Fragen und Anforderungen sammeln und im Rahmen von Best Practices Leistungsindikatoren zur Messung von Verbesserungen definieren. Beispiele für primäre Forschungsfragen können sein:

- Wie ist das menschliche Wohlergehen mit der Biodiversität verknüpft?
- Wie beeinflussen Mikroorganismen die Produktivität in der Landwirtschaft, die Gesundheit, neuronale oder immunologische Prozesse und die Biodiversität?
- Wie könnte die Vernetzung bestehender Datenbanken verbessert werden, um die allgemeine Verbreitung und spezifische Ausbrüche von Infektionskrankheiten bei Mensch und Tier schneller aufzuklären?
- Wie wirkt sich der Klimawandel auf die biologische Vielfalt oder das menschliche Wohlbefinden aus?

Des Weiteren wird die Forschung von einer Vielzahl von kleineren Fachcommunities durchgeführt, die ein breites Spektrum unterschiedlicher und heterogener Daten nutzen. Diese umfassen die biologische Vielfalt, die Molekularbiologie bis hin zu Medizin, Landwirtschaft und Ernährung. Die daraus erwachsene große Anzahl vorhandener Datenquellen, darunter Fakten-Datenbanken (z.B. UniProtKb, BRENDA, SILVA, PANGAEA, TCGA), Repositorien für Biodiversitätsdaten (gemeinsam z.T. über GBIF- und GFBio-Portaldienste zugänglich), Genomik- und Proteomikdaten-Repositorien (ENA, ArrayExpress, PRIDE) und zunehmend auch Bilddaten, tragen zu einer effizienten und innovativen Forschung in den verschiedenen Forschungsbereichen bei.

Die Gewährleistung der Interoperabilität all dieser Datenrepositorien fördert die Möglichkeiten der Analyse und Wiederverwendung der experimentellen Daten. Bei der Analyse der eigenen Daten benötigen Forschende möglicherweise Zugang zu einer Cloud-Infrastruktur mit den Referenzdatenquellen und extrahieren entweder die relevantesten Informationen für lokale Analysen oder laden die eigenen Daten zur cloud-basierten Verarbeitung hoch. Diesen Aufgaben wird sich NFDI4Life Umbrella widmen.

Maßnahmen und Services (1) NFDI4Life Umbrella unterstützt und koordiniert die Netzwerkaktivitäten der einzelnen (geförderten und noch nicht geförderten) lebenswissenschaftlichen Konsortien, so dass diese Aktivitäten die Konsortien entlasten. Diese können sich besser auf die besonderen Bedürfnisse ihrer jeweiligen Fachcommunities konzentrieren. (2) NFDI4Life Umbrella unterstützt und vertritt die Konsortienengemeinschaft als Organ auf politischer Ebene. (3) NFDI4Life Umbrella kartiert und überwacht kontinuierlich die Landschaft der Lebenswissenschaften im Hinblick auf bestehende rechtliche und regulatorische Widersprüche und soll so Entscheidungsträgern und Geldgebern bei der Suche nach effizienten Lösungen helfen. (4) Da sich die infrastrukturelle Landschaft in den Lebenswissenschaften dynamisch entwickelt, ist es wichtig, rechtzeitig eine übergreifende Koordinierungsstelle zu schaffen, die Parallelentwicklungen verhindert und die Standardisierungsprozesse der verschiedenen NFDIs in den Lebenswissenschaften harmonisiert.

Beispiele für konkrete Maßnahmen sind die Schaffung domänenübergreifender Metadatenstandards für die Lebenswissenschaften, die Förderung von Bildung, Ausbildung und beruflicher Entwicklung durch Workshops, Webinare, Summer schools, Graduiertenschule und ein Informationsportal, die Förderung der Datenwissenschaft und Fähigkeiten im Forschungsdatenmanagement als Teil der aktuellen Lehrpläne, die Einführung von

Leistungsindikatoren und Benchmarking, die Konsolidierung rechtlicher Rahmenwerke, Empfehlungen für Werkzeuge und Methoden (nach Beurteilung ihrer Qualität), ein Terminologie-Nachschlagedienst zusammen mit Beratungsdienstleistungen, die Ermittlung von Diensten und Instrumenten, die auf andere Bereiche der Lebenswissenschaften übertragen werden können, z.B. Meta-Meta-Dienst für European Open Science Cloud (EOSC), Beratung und Lösungen für elektronische Laborbücher (ELN) als erster und zentraler Teil einer Infrastruktur zur Verwaltung von Forschungsdaten, Dienste für die Zuweisung persistenter Identifikatoren, Dienste zur Erstellung von Datenmanagementplänen, langfristige Datenerhaltung und Datenarchivierung, IT/Cloud-Entwicklung und Infrastrukturen (zum Beispiel de.NBI), Text- und Data-Mining-Werkzeuge für Datenannotation oder Informationsextraktion aus Textressourcen, Erstellung und Verbreitung von Richtlinien sowie die Festlegung ethischer und technischer Rahmenbedingungen für den Umgang mit sensiblen Daten.

Datenarten NFDI4Life-Umbrella adressiert übergreifende Metadaten als Daten. Das bedeutet, dass Forschende die Möglichkeit erhalten, ihre Metadaten (terminologische und ontologischen Quellen), als Input zu liefern, diese auf einer zentralen Seite zu hosten (welche die Interoperabilität und FAIR-Parameter analysiert) und die Metadaten in verschiedenen (transformierten) Datenstandards wieder ausliefern zu lassen.

Querschnittsthemen NFDI4Life Umbrella sieht sich selber als Querschnittskonsortium für die spezifischen Belange der Lebenswissenschaften. Die Initiative adressiert die in den vorstehenden Abschnitten aufgeführten Aufgaben und referenziert darüber hinaus auf die "Leipzig-Berlin-Erklärung zu NFDI-Querschnittsthemen der Infrastrukturentwicklung" (<http://doi.org/10.5281/zenodo.3895209>).

Vernetzungsaktivitäten Für NFDI4Life Umbrella sind Vernetzungsaktivitäten wesentlich, um Sichtbarkeit aufzubauen und zu erhöhen und mit einer Stimme für die Lebenswissenschaften sprechen zu können. Daher wird sich das Konsortium sowohl auf nationaler als auch internationaler Ebene um eine starke Vernetzung bemühen. Auf nationaler Ebene erfolgt neben der Verbindung zu anderen NFDI-Konsortien beispielsweise ein Austausch mit deutschen wissenschaftlichen Sammlungen und Biobanken, mit den Präsidenten der Wissenschaftsorganisationen und der Allianz der Wissenschaftsorganisationen sowie mit den Hochschulen und der Hochschulrektorenkonferenz zur koordinierten Entwicklung von Lehre und Curricula.

Auf internationaler Ebene wird z.B. die Verbindung zur European Open Science Cloud (EOSC), insbesondere EOSC-Life, der Research Data Alliance (RDA), den Europäischen FAIR-Initiativen und -Infrastrukturen: Go Fair, FAIRDOM, FAIRplus, usw., zu ELIXIR (de.NBI ist der deutsche Knotenpunkt von ELIXIR) und zu ESFRI Biologische und medizinische Forschungsinfrastrukturen, wie BBMRI-ERIC (Biobanking and BioMolecular resources Research Infrastructure) und andere Initiativen unter dem Dach von CORBEL, angestrebt.

Erwartung an die NFDI-Konferenz NFDI4Life Umbrella erhofft sich eine Klärung zum Umgang mit Querschnittsthemen und ggf. Konsortien, die diese Themen adressieren, da NFDI4Life Umbrella sich selber als fachlich bezogenes Querschnittskonsortium sieht.

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

Co-Sprecher/in	Zugehörige Institution
Iris Pigeot Direktorin Sprecherin NFDI4Health pigeot@leibniz-bips.de	Leibniz-Institut für Präventionsforschung und Epidemiologie - BIPS Achterstraße 30 28359 Bremen
Alfred Pühler Koordinator de.NBI-Netzwerk, Leiter ELIXIR-DE puehler@CeBiTec.Uni-Bielefeld.DE	Center for Biotechnology - CeBiTec Universitätsstraße 25 33615 Bielefeld
Wolfgang Müller Gruppenleiter und Institutssprecher wolfgang.mueller@h-its.org	Heidelberger Institut für Theoretische Studien - HITS gGmbH Schloss-Wolfsbrunnenweg 35 69118 Heidelberg
Ulrich Lang Leiter Rechenzentrum lang@uni-koeln.de	Rechenzentrum der Universität zu Köln Weyertal 121 50931 Köln
Hubertus Neuhausen Direktor direktor@ub.uni-koeln.de	Universitäts- und Stadtbibliothek, Universität zu Köln Universitätsstraße 33 50931 Köln
Uwe Heinrich Leiter AG Geodaten uheinrich@zalf.de	Leibniz-Zentrum für Agrarlandschaftsforschung (ZALF) e. V. Eberswalder Straße 84 15374 Müncheberg
Frank Oliver Glöckner Leiter PANGAEA Sprecher NFDI4BioDiversity frank.oliver.gloeckner@awi.de	MARUM - Center for Marine Environmental Sciences University of Bremen Leobener Str. 8 D-28359 Bremen Alfred Wegener Institute - Helmholtz Center for Polar- and Marine Research Am Handelshafen 12 27570 Bremerhaven
Sebastian C. Semler Geschäftsführer sebastian.semmler@tmf-ev.de	TMF - Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V Charlottenstraße 42/Ecke Dorotheenstraße 10117 Berlin

Co-Sprecher/in	Zugehörige Institution
<p>Oya Beyan Stellvertr. Leiterin Abt. Knowledge Pipelines beyan@fit.fraunhofer.de</p>	<p>Fraunhofer Institute for Applied Information Technology Schloss Birlinghoven, Konrad-Adenauer-Straße 53754 Sankt Augustin</p>
<p>Linus Grabenhenrich Leitung Forschungsdatenmanagement grabenhenrichl@rki.de</p>	<p>Robert Koch-Institut Nordufer 20 13353 Berlin</p>
<p>Dr. Uwe Scholz Leiter Forschungsgruppe Bioinformatik und Informationstechnologie; Koordinator Biodiversitätsinformatik scholz@ipk-gatersleben.de</p>	<p>Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK) Gatersleben Corrensstraße 3 06466 Seeland</p>
<p>Ulrich Sax Leitung und Koordination Institut für Medizinische Informatik ulrich.sax@med.uni-goettingen.de</p>	<p>University Medical Center Göttingen Robert-Koch-Str. 40 37075 Göttingen</p>
<p>Lothar Kreienbrock Institutsdirektor Institut für Biometrie, Epidemiologie und Informationsverarbeitung lothar.kreienbrock@tiho-hannover.de</p> <p>Sabine Leonhard-Marek Leitung Bibliothek sabine.leonhard-marek@tiho-hannover.de</p>	<p>Tierärztliche Hochschule Hannover Bünteweg 2 30559 Hannover</p>
<p>Christian Busse Projektleiter Sprecher NFDI4AIRR christian.busse@dkfz-heidelberg.de</p>	<p>Deutsches Krebsforschungszentrum (DKFZ) Im Neuenheimer Feld 280 69120 Heidelberg</p>
<p>Prof. Dr. Björn Usadel Direktor Institut für biologische Datenwissenschaften, IBG-4 Sprecher DataPlant b.usadel@fz-juelich.de</p>	<p>Forschungszentrum Jülich, Wilhelm Johnen Strasse 52428 Jülich</p>
<p>Dr. Jens Krüger Gruppenleiter High Performance und Cloud Computing Sprecher DataPlant jens.krueger@uni-tuebingen.de</p>	<p>Eberhard Karls Universität Tübingen Geschwister-Scholl-Platz 72074 Tübingen</p>

Co-Sprecher/in	Zugehörige Institution
Prof. Dr. Steve Hoffmann Leiter Computational Biology Sprecher MOPED steve.hoffmann@leibniz-fli.de	Leibniz-Institut für Altersforschung Beutenbergstrasse 11 07745 Jena
Prof. Dr. Christoph Steinbeck Prof. für Analytische Chemie, Chemieinformatik und Chemometrics Sprecher NFDI4Chem christoph.steinbeck@uni-jena.de	Friedrich-Schiller-University Jena Fürstengraben 1 07743 Jena