

## Non-binding letter of intent NFDI4Life Umbrella

### 1 Binding letter of intent as advance notification or non-binding letter of intent

<input type="checkbox"/>	Binding letter of intent (required as advance notification for proposals in 2020)
<input checked="" type="checkbox"/>	Non-binding letter of intent (anticipated submission in 2021)

### 2 Formal details

- Planned name of the consortium  
NFDI4Life Umbrella research data management infrastructure for life sciences
- Acronym of the planned consortium  
NFDI4Life Umbrella
- Applicant institution  
ZB MED Information Centre Life Sciences, Gleuelerstr. 60, 50931 Cologne  
Head: Prof. Dr. Dietrich Rebholz-Schuhmann
- Spokesperson  
Prof. Dr. Dietrich Rebholz-Schuhmann, ZB MED Information Centre Life Sciences,  
[rebholz-schuhmann@zbmed.de](mailto:rebholz-schuhmann@zbmed.de)
- Co-applicant institution  
[Name and address]  
[Name of the head of the co-applicant institution]
- Co-spokesperson  
[Name, e-mail address, institutional affiliation]

See table below

<p><b>Deutsche Forschungsgemeinschaft</b> Kennedyallee 40 · 53175 Bonn, Germany · Postal address: 53170 Bonn, Germany Tel.: + 49 228 885-1 · Fax: + 49 228 885-2777 · <a href="mailto:postmaster@dfg.de">postmaster@dfg.de</a> · <a href="http://www.dfg.de">www.dfg.de</a></p>	
---	--

Co-spokesperson	Co-applicant institution
<p>Iris Pigeot                      Director                      Spokesperson NFDI4Health                      pigeot@leibniz-bips.de</p>	<p>Leibniz-Institut für                      Präventionsforschung und                      Epidemiologie - BIPS                      Achterstraße 30                      28359 Bremen</p>
<p>Alfred Pühler                      Coordinator de.NBI-Netzwerk, Head of                      ELIXIR-DE                      puehler@CeBiTec.Uni-Bielefeld.DE</p>	<p>Center for Biotechnology - CeBiTec                      Universitätsstraße 25                      33615 Bielefeld</p>
<p>Wolfgang Müller                      Spokesperson and group leader                      wolfgang.mueller@h-its.org</p>	<p>Heidelberger Institut für Theoretische                      Studien - HITS gGmbH                      Schloss-Wolfsbrunnenweg 35                      69118 Heidelberg</p>
<p>Ulrich Lang                      Head of IT-Center                      lang@uni-koeln.de</p>	<p>Rechenzentrum der Universität zu Köln                      Weyertal 121                      50931 Köln</p>
<p>Hubertus Neuhausen                      Director                      direktor@ub.uni-koeln.de</p>	<p>Universitäts- und Stadtbibliothek,                      Universität zu Köln                      Universitätsstraße 33                      50931 Köln</p>
<p>Uwe Heinrich                      Head of AG Geodaten                      uheinrich@zalf.de</p>	<p>Leibniz-Zentrum für                      Agrarlandschaftsforschung (ZALF) e. V.                      Eberswalder Straße 84                      15374 Müncheberg</p>

Co-spokesperson	Co-applicant institution
<p>Frank Oliver Glöckner                      Head of PANGAEA                      Spokesperson NFDI4BioDiversity                      frank.oliver.gloeckner@awi.de</p>	<p>MARUM - Center for Marine                      Environmental Sciences                      University of Bremen                      Leo-bener Str. 8                      D-28359 Bre-men</p> <p>Alfred Wegener Institute - Helmholtz                      Center for Polar- and Marine Research                      Am Handelshafen 12                      27570 Bremerhaven</p>
<p>Sebastian C. Semler                      Managing director                      sebastian.semmler@tmf-ev.de</p>	<p>TMF - Technologie- und                      Methodenplattform für die vernetzte                      medizinische Forschung e.V                      Charlottenstraße 42/Ecke                      Dorotheenstraße                      10117 Berlin</p>
<p>Oya Beyan                      Deputy head department 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                      Head of research data management                      grabenhenrich@rki.de</p>	<p>Robert Koch-Institut                      Nordufer 20                      13353 Berlin</p>
<p>Dr. Uwe Scholz                      Head of Bioinformatic and Information                      Technology; coordinator Biodiversity                      Informatics                      scholz@ipk-gatersleben.de</p>	<p>Leibniz-Institut für Pflanzengenetik und                      Kulturpflanzenforschung (IPK)                      Gatersleben                      Corrensstraße 3                      06466 Seeland</p>
<p>Ulrich Sax                      Head of Institute of medicine                      informatics                      ulrich.sax@med.uni-goettingen.de</p>	<p>University Medical Center Göttingen                      Robert-Koch-Str. 40                      37075 Göttingen</p>

Co-spokesperson	Co-applicant institution
<p>Lothar Kreienbrock                      Director Institut für Biometrie,                      Epidemiologie und                      Informationsverarbeitung                      lothar.kreienbrock@tiho-hannover.de</p> <p>Sabine Leonhard-Marek                      Head of library                      sabine.leonhard-marek@tiho-hannover.de</p>	<p>Tierärztliche Hochschule Hannover                      Bünteweg 2                      30559 Hannover</p>
<p>Christian Busse                      Project leader Digital Immunology                      Spokesperson NFDI4Immuno                      christian.busse@dkfz-heidelberg.de</p>	<p>Deutsches Krebsforschungszentrum                      (DKFZ)                      Im Neuenheimer Feld 280                      69120 Heidelberg</p>
<p>Prof. Dr. Björn Usadel                      Director Institut für biologische                      Datenwissenschaften, IBG-4                      Spokesperson DataPlant                      b.usadel@fz-juelich.de</p>	<p>Forschungszentrum Jülich,                      Wilhelm Johnen Strasse                      52428 Jülich</p>
<p>Dr. Jens Krüger                      Group leader High Performance und                      Cloud Computing                      Spokesperson DataPlant                      jens.krueger@uni-tuebingen.de</p>	<p>Eberhard Karls Universität Tübingen                      Geschwister-Scholl-Platz                      72074 Tübingen</p>
<p>Prof. Dr. Steve Hoffmann                      Head Computational Biology                      Spokesperson MOPED                      steve.hoffmann@leibniz-fli.de</p>	<p>Leibniz-Institut für Alternsforschung                      Beutenbergstrasse 11                      07745 Jena</p>
<p>Prof. Dr. Christoph Steinbeck                      Prof. for Analytical Chemistry,                      informatics and chemometrics                      Spokesperson NFDI4Chem                      christoph.steinbeck@uni-jena.de</p>	<p>Friedrich-Schiller-University Jena                      Fürstengraben 1                      07743 Jena</p>

- Participant  
[Name, institutional affiliation]

to announce later in case of a binding letter of intent in 2021

### 3 Objectives, work programme and research environment

- Research area of the proposed consortium (according to the DFG classification system:

[https://www.dfg.de/download/pdf/dfg\\_im\\_profil/gremien/fachkollegien/amtsperiode\\_2020\\_2023/fachsystematik\\_fachkollegien\\_amtsperiode\\_2020-2023.pdf](https://www.dfg.de/download/pdf/dfg_im_profil/gremien/fachkollegien/amtsperiode_2020_2023/fachsystematik_fachkollegien_amtsperiode_2020-2023.pdf) )

[Multiple review boards (Fachkollegien) or subject areas (Fächer) may be listed.]

2 Life Sciences (201, 202, 203, 204, 205, 206, 207)

- Concise summary of the planned consortium's main objectives and task areas

The NFDI4Life Umbrella addresses all overarching needs of research data management in the life sciences with all its subdomains.

The main objectives of the NFDI4Life Umbrella are:

- Cross-subdomain interoperability of data sources in the life science domain in general: Ensuring FAIR data across all life science domain NFDI consortia,
- Standardization of cross-subdomain used data and processes, especially of relevant subdomain-overarching metadata structures specific for the life sciences, and harmonization of standards, as well as coordination of the processes across NFDI consortia in the whole life science domain (according to demands),
- Sharing of IT solutions and in particular cloud-based IT solutions developed and provided for the purpose of joint use by all life science domain NFDI consortia,
- Out-reach to the whole spectrum of life science research communities at large and across subdomain-specific consortia through concerted education and training, as well as international representation and visibility of the German life science data community in Europe: an organ of the life science NFDIs to politics, funders, the Allianz der Wissenschaftsorganisationen, other NFDI consortia, and international organizations such as Go FAIR, RDA and EOSC,
- NFDI4Life Umbrella aims to be a consolidated and strong voice of the German life science communities in national, European and international debates on policies, regulations or standards,
- Provide solutions for specific life science aspects in generic tasks such as reputation, cultural change, legal advice/opinions (e.g. to address questions around GDPR that are specific for Life Science) or policy making, that can be better addressed in a life science

**Deutsche Forschungsgemeinschaft**

Kennedyallee 40 · 53175 Bonn, Germany · Postal address: 53170 Bonn, Germany

Tel.: + 49 228 885-1 · Fax: + 49 228 885-2777 · [postmaster@dfg.de](mailto:postmaster@dfg.de) · [www.dfg.de](http://www.dfg.de)

umbrella concept but not addressed in the same way by subdomain-specific NFDI consortia,

- Linking together the existing initiatives with relation to research data management from scientific communities and information infrastructures in the life sciences at large.

The main task areas are divided in two blocks:

## 1. Networking and coordination

The NFDI4Life Umbrella consortium is meant to enable cross-cutting support for NFDI consortia that focus on their own subdomain communities in terms of research data management, but would require vital support for the integration of their community data into the full range of all communities across the life sciences at large. It has been already demonstrated that data from different parts of the life science domain (see molecular biology and medicine) is relevant beyond the specific limits of a given subdomain. However each consortium in one particular subdomain of the life sciences encounters overheads in achieving interoperability with the other subdomains, and all consortia in the life sciences at large will encounter partially similar overheads. Reducing these overheads and aligning all the different types of data can be achieved through subdomain-overarching centralized services (with contributions from the different consortia) and will lead to benefits in the different consortia, e.g. through extended use of the research data across all consortia. Such infrastructures have been envisaged from the ESFRI network ELIXIR, but have only (partially) instantiated in the life science domain (in contrast to other scientific domains).

Specific tasks that will be addressed from the NFDI4Life Umbrella consortium to achieve networking and coordination benefits comprise the following: administering and fostering contacts between the consortia; enabling sustained coordination of joined services, interfaces and standards, e.g. dissemination of terminologies; organization of workshops jointly for members from the life science consortia; active exchange and alignment of guidelines and policies including provision of necessary repository support; collecting the requirements and contributions of the consortia concerning cross cutting topics, sources and solutions; and positioning the NFDI consortia, the semantic resources and infrastructures in the life science community (“outreach”).

## 2. Infrastructure and services

Data and services are highly intertwined in the life science research domain at large and the high demand for data analytics required (e.g., for genomics data) initiated the development of cloud infrastructures for public use in the research community. As a result, a distributed cloud infrastructure (see de.NBI <https://www.denbi.de>) is available in combination with the know-how of data management and data analytics to set the current and future standards for the (re-)use of research data in the research community.

On the other side, this infrastructure setup sets the premises for further growth, further alignment of existing data resources, further reuse of data and tools in a cloud-based infrastructure and forms the central unit for scaling up the overall performance of research communities on a national level.

Specific tasks that would be addressed by the NFDI4Life Umbrella consortium are concerned with the overarching semantic standardization across all subdomains in the life sciences, which

<p><b>Deutsche Forschungsgemeinschaft</b>  Kennedyallee 40 · 53175 Bonn, Germany · Postal address: 53170 Bonn, Germany  Tel.: + 49 228 885-1 · Fax: + 49 228 885-2777 · <a href="mailto:postmaster@dfg.de">postmaster@dfg.de</a> · <a href="http://www.dfg.de">www.dfg.de</a></p>	
---	--

comprises the support and supervision of the terminology development (in particular the semi-automatic alignment of terminologies across consortia), the mapping of semantic resources from different subdomains for better reuse of resources, the automatic adjustment of new semantic resources to existing ones, and the development and provision of the infrastructure for using, access and maintenance of terminologies, ontologies and other semantic resources. Further tasks would be concerned with the indexing, dissemination of and access to metadata, the advanced search for (meta) data (and the ongoing development of advanced searches), and the integration of data with the scientific literature through semantic solutions.

The before-mentioned solutions are geared to serve the scientist to best connect the own data through the existing IT infrastructure with the whole pool of data and literature in the life sciences at large in an established cloud infrastructure, and by these means to drive the academic process of the scientist providing own data, or even better own data in combination with the publication about the data and research, both as active and reusable elements in the cloud-based infrastructure.

- Brief description of the proposed use of existing infrastructures, tools and services that are essential in order to fulfil the planned consortium's objectives

The German life sciences already feature well advanced infrastructures and networks, like the German National Cohort (NAKO Health Study, <https://nako.de>), the German Network for Plant Phenotyping (DPPN) or 'Soil as a sustainable resource for the bioeconomy' (BonaRes), and together with European partners the infrastructure FAIRDOM (<https://fair-dom.org>) that supports researchers in the life sciences with data management since more than 10 years. Furthermore the DFG-funded, multidisciplinary consortium German Federation for Biological Data (GFBio, <http://www.gfbio.org>) is an existing infrastructure, that has undergone a six year formation process regarding infrastructure and community-building.

Another actor is the platform for Technology, Methods, and Infrastructure for Networked Medical Research (TMF e.V., <http://www.tmf-ev.de>) which is currently coordinating the Medical Informatics Initiative (MII) together with Medizinischer Fakultätentag (MFT) and Verband der Universitätsklinika Deutschlands (VUD). Its vision is the development and deployment of expert opinions, generic concepts, specimen texts, and IT applications, as well as training and consultation to strengthen the quality and efficiency of medical research and to clarify the legal and ethical foundations for performing medical research.

Additionally, the handling, analysis and storage of enormous amounts of data is a challenging issue across all subdomains in state-of-the-art life science research. Hence, an appropriate IT infrastructure is crucial for performing big data analyses and ensuring secure data access and storage. The cloud infrastructure of the German Network for Bioinformatics Infrastructure (de.NBI Cloud, <https://www.denbi.de/cloud>) has been established over the last years to enable integrative analyses for the entire life sciences community in Germany and the efficient use of data in research and application.

The use of the de.NBI cloud plays an important role in already selected NFDI projects, such as NFDI4BioDiversity and GHGA. Furthermore, the de.NBI cloud is also heavily involved in the planned NFDI4Microbiota project.

**Deutsche Forschungsgemeinschaft**

Kennedyallee 40 · 53175 Bonn, Germany · Postal address: 53170 Bonn, Germany  
Tel.: + 49 228 885-1 · Fax: + 49 228 885-2777 · [postmaster@dfg.de](mailto:postmaster@dfg.de) · [www.dfg.de](http://www.dfg.de)

ZB MED has the national and institutional task to provide access to scientific literature and data in the life science domain. As part of this task, it actively covers tasks such as training in semantic resources (data and metadata management, e.g., for librarians and data scientists), provides an IT infrastructure for access to the scientific literature, to metadata information and to a knowledge environment in the life science domain (increasingly in combination with the cloud infrastructure of the de.NBI site at the University of Bielefeld), and drives research in the efficient use of semantic resources in the domain of text and data mining for the life sciences. Although ZB MED combines different relevant competencies, it heavily relies on research and service partnerships in all domains to support a research infrastructure at scale according to the contributions of ZB MED's partners and ZB MED in itself.

- Interfaces to other proposed NFDI consortia: brief description of existing agreements for collaboration and/or plans for future collaboration

Other proposed discipline specific life science consortia agree to support and contribute to the NFDI4Life Umbrella as coordinative council:

- NFDI4BioDiversity
- NFDI4Health
- NFDI4Microbiota
- NFDI Neuroscience
- NFDI4Immuno
- NFDI4Agri
- German Genome-Phenome Archive (GHGA)
- DataPLANT
- NFDI4Chem

Furthermore the collaboration with other cross-cutting consortia is planned to match the needs of the life science with the special aspects of their topic:

- CompeNDI (Competencies for NFDI) for the topic of education and training
- NFDI4RSE for the topic of research software engineering
- NFDI4DataScience for bridging technological solutions dealing with research data at scale

#### 4 Cross-cutting topics

- Please identify cross-cutting topics that are relevant for your consortium and that need to be designed and developed by several or all NFDI consortia.

The following cross-cutting topics have been identified from the NFDI4Life Umbrella Consortium (according to a joint workshop in Oct 2018).

Tasks closely concerned with the data from the different subdomain-specific NFDI consortia in

<p><b>Deutsche Forschungsgemeinschaft</b> Kennedyallee 40 · 53175 Bonn, Germany · Postal address: 53170 Bonn, Germany Tel.: + 49 228 885-1 · Fax: + 49 228 885-2777 · postmaster@dfg.de · www.dfg.de</p>	
--	--

the life sciences comprises the data management planning (and related planning of tools and solutions), namely the Research Data Management Organizer for the life sciences (RDMO4Life <https://rdmo.publisso.de/>) solution as an approach to work towards a standardized way for the ingestion of high quality research data, as well as the use of shared life science specific, but subdomain-overarching terminological and metadata standards, resources, tools and solutions to achieve semantic interoperability across consortia. This work leads to the task for long term preservation of digital data, which includes maintenance, access and reuse of the data and research objects. One central piece in this data provision approach form the services for persistent identifiers (PID) that are essential to have the data in the public through the data infrastructure.

As part of the data preservation and data archiving tasks, the legal aspects for data provision have to be kept in mind and addressed through experts in this domain, who would ensure that data is openly available, the licence agreements fit the needs of the researcher and the public alike, and long-term provision is guaranteed from the submission day onwards.

Other specific tasks would address the research community in the sense that they benefit researchers who want to contribute their own data. This contribution includes specific conceptualization, planning, development and roll-out of solutions that would drive the reputation credits of the scientist and would contribute to the culture building for research data management and publication benefits. In essence, delivering own data into a research data management infrastructure should promote the visibility and the scientific credits of the scientist (see PIDs and adjacent publication for the research data), should enable the scientists to use the open data repository to increase the academic benefits for the researcher, and should serve as a hub for science overall.

All the tasks above have to be accompanied with education and training to achieve a shared understanding of the data infrastructure, where the shared understanding is based on already existing and used semantic resources, tools and solutions in the life sciences, but would also grow with the set of new solutions from the different consortia in the whole life science domain. The initiative also refers to the "Leipzig-Berlin Declaration on NFDI Cross-cutting Issues of Infrastructure Development" (<http://doi.org/10.5281/zenodo.3895209>).

- Please indicate which of these cross-cutting topics your consortium could contribute to and how.

Due to the structure of NFDI4Life Umbrella spanning all life science subdomains, the consortium can contribute as a unified voice of the life sciences to all of the above mentioned cross-cutting topics. There will already be an interdisciplinary alignment for these topics that can be rolled out to other disciplines or tested for the whole NFDI.