

NFDI4Microbiota - Letter of Intent for the National Research Data Infrastructure (NFDI)

2020-08-15

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

- Binding letter of intent (required as advance notification for proposals in 2020)
- Non-binding letter of intent (anticipated submission in 2021)

2. Formal details

- *Planned name of the consortium:* National Research Data Infrastructure for Microbiota Research
- *Acronym of the planned consortium:* NFDI4Microbiota
- *Applicant institution:*

ZB MED - Information Centre for Life Sciences

Gleueler Straße 60

50931 Cologne (Köln)

Scientific Director: Prof. Dr. Dietrich Rebholz-Schuhman

- *Spokesperson:*

Prof. Dr. Konrad Förstner, foerstner@zbmed.de, ZB MED - Information Centre for Life Sciences

- *Co-applicant institution:*

German Collection of Microorganisms and Cell Cultures (DSMZ)

Inhoffenstraße 7B

38124 Braunschweig

Director: Prof. Dr. Jörg Overmann

European Molecular Biology Laboratory (EMBL)

Meyerhofstraße 1

69117 Heidelberg

Director General: Prof. Edith Heard (PhD)

Helmholtz Centre for Infection Research (HZI)

Inhoffenstraße 7

38124 Braunschweig

Scientific Director: Prof. Dr. Dirk Heinz

Administrative Director: Silke Tannapfel

Bielefeld Institute for Bioinformatics Infrastructure (BIBI), Bielefeld University

Universitätsstraße 25

33615 Bielefeld

Rector: Prof. Dr. Gerhard Sagerer

Friedrich Schiller University Jena (FSU Jena)

Fuerstengraben 1

07743 Jena

President: Prof. Dr. Walter Rosenthal

Helmholtz Centre for Environmental Research (UFZ)

Permoserstraße 15

04318 Leipzig

Scientific Managing Director: Prof. Dr. Georg Teutsch

Administrative Director: Dr. Sabine König

Justus-Liebig-University Giessen (JLU Giessen)

Ludwigstraße 23

35390 Gießen

President: Prof. Dr. Joybrato Mukherjee

Philipps-Universität Marburg (UMR)

Biegenstraße 10

35037 Marburg

President: Prof. Dr. Katharina Krause

RWTH Aachen University (RWTH)

Templergraben 55

52062 Aachen

Rector: Prof. Dr. Dr. Ulrich Rüdiger

- *Co-spokesperson*

Prof. Dr. Peer Bork, bork@embl.de, European Molecular Biology Laboratory (EMBL)

Prof. Dr. Alice McHardy, Alice.McHardy@helmholtz-hzi.de, Helmholtz Centre for Infection Research (HZI)

Prof. Dr. Jörg Overmann, joerg.overmann@dsmz.de, German Collection of Microorganisms and Cell Cultures (DSMZ)

Prof. Dr. Jens Stoye, jens.stoye@uni-bielefeld.de, Bielefeld Institute for Bioinformatics Infrastructure (BIBI), Bielefeld University

Prof. Dr. Alexander Sczyrba, asczyrba@cebitec.uni-bielefeld.de, Bielefeld Institute for Bioinformatics Infrastructure (BIBI), Bielefeld University

Prof. Dr. Manja Marz, manja@uni-jena.de, RNA Bioinformatics & High-Throughput Analysis, Friedrich Schiller University Jena (FSU Jena)

Dr. Ulisses Nunes da Rocha, ulisses.rocha@ufz.de, Department of Environmental Microbiology, Helmholtz Centre for Environmental Research (UFZ)

Prof. Dr. Alexander Goesmann, alexander.goesmann@computational.bio.uni-giessen.de, Bioinformatics & Systems Biology, Justus-Liebig-University Giessen (JLU Giessen)

Prof. Dr. Anke Becker, anke.becker@synmikro.uni-marburg.de, SYNMIKRO, Philipps-Universität Marburg (UMR)

Prof. Dr. Thomas Clavel, tclavel@ukaachen.de, Medical Microbiology, RWTH Aachen University (RWTH)

- *Participant*

- Prof. Dr. Anita Marchfelder, SPP 2141 – CRISPR-Cas functions beyond defense, Ulm University
- Prof. Dr. Ruth Schmitz-Streit, SPP2002 – Small Proteins in Prokaryotes, an Unexplored World, University of Kiel
- Prof. Dr. Julia Frunzke, SPP 2330 - New concepts in prokaryotic virus-host interaction – from single cells to microbial communities, Jülich Research Centre
- Prof. Dr. Dirk Haller, SFB 1371 - Microbiome Signatures, Technical University of Munich
- Prof. Dr. Stephan Becker, SFB 1021 - RNA viruses: RNA metabolism, host response and pathogenesis, Philipps-Universität Marburg (UMR)
- Prof. Dr. Axel A. Brakhage, SFB/TR 124 – Pathogenic fungi and their human host: Networks of interaction, Friedrich Schiller University Jena (FSU Jena)
- Prof. Dr. Axel A. Brakhage, EXE 2051 - Cluster of Excellence Balance of the Microverse, Friedrich Schiller University Jena (FSU Jena)
- Prof. Dr. Folker Meyer, Institute for Artificial Intelligence, Essen University Hospital
- Prof. Dr. Gerd Geisslinger, Fraunhofer Cluster of Excellence for Immune-Mediated Diseases
- Prof. Dr. Kai Papenfort, Chair of General Microbiology, Friedrich Schiller University of Jena - Institute of Microbiology
- Dr. Ralf Möller, Space Microbiology Research Group, German Aerospace Center (DLR)
- Prof. Dr. Jörg Soppa, Department of Biology and Genetics of Prokaryotes, Goethe University Frankfurt
- Prof. Dr. Alexander Diltthey, Institute of Medical Microbiology and Hospital Hygiene, Heinrich Heine University Düsseldorf
- Dr. Alexander Bartholomäus, Section 3.7: Geomicrobiology, Helmholtz Centre Potsdam - GFZ German Research Centre for Geosciences
- Prof. Dr. Jörg Vogel, RNA Biology of Bacterial Infections, Helmholtz Institute for RNA-based Infection Research (HIRI)
- Prof. Dr. Michael Schlöter, Research Unit for Comparative Microbiome Analysis, Helmholtz Zentrum München German Research Center for Environmental Health
- Dr. Tilman Klassert, Host Septomics Research Group, Jena University Hospital - Centre for Innovation Competence (ZIK) Septomics
- Prof. Dr. Michael Bott, Institute of Bio- and Geosciences – Biotechnology (IBG-1), Jülich Research Centre
- Dr. Markus Engstler, Cell function and organ development, Julius Maximilian University of Würzburg
- Prof. Dr. Thomas Dandekar, Department of Bioinformatics, Julius Maximilian University of Würzburg
- Prof. Dr. Cynthia Sharma, Institute for Molecular Infection Biology (IMIB), Julius Maximilian University of Würzburg
- Prof. Dr. Lars Dölken, Institute for Virology and Immunobiology, Julius Maximilian University of Würzburg
- Prof. Dr. Cynthia Sharma, Research Center for Infectious Diseases (ZINF), Julius Maximilian University of Würzburg
- Prof. Dr. Gabriele Klug, Institute of Microbiology and Molecular Biology, Justus-Liebig-Universität Giessen

- Prof. Dr. Linda Falgenhauer, Institute of Hygiene and Environmental Medicine, Justus-Liebig-University Giessen (JLU Giessen)
- Prof. Dr. Trinad Chakraborty, Institute of Medical Microbiology, Justus-Liebig-University Giessen (JLU Giessen)
- Prof. Dr. John Ziebuhr, Institute of Medical Virology, Justus-Liebig-University Giessen (JLU Giessen)
- Prof. Dr. Johannes Gescher, Institute for Applied Biology (IAB), Karlsruhe Institute of Technology
- Prof. Dr. Axel A. Brakhage, Infection Biology and Molecular Biotechnology, Leibniz Institute for Natural Product Research and Infection Biology – Hans Knöll Institute (HKI)
- Assoc. Prof. Dr. Gianni Panagiotou, Systems Biology and Bioinformatics, Leibniz Institute for Natural Product Research and Infection Biology – Hans Knöll Institute (HKI)
- Prof. Dr. Steve Hoffmann, Computational and Systems Biology of Aging, Leibniz Institute on Aging – Fritz Lipmann Institute (FLI)
- Prof. Dr. Hans-Peter Grossart, Aquatic microbial ecology, Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB)
- Prof. Dr. Nicolai Siegel, Biomedical Center Munich, Ludwig Maximilian University of Munich
- Prof. Dr. Nicolai Siegel, Experimental Parasitology, Ludwig Maximilian University of Munich
- Prof. Dr. Timo Niedermeyer, RG Pharmacognosy, Martin Luther University Halle-Wittenberg (MLU)
- Dr. Sofia Forslund, Host-microbiome factors in cardiovascular disease, Max Delbrück Center for Molecular Medicine
- Prof. Dr. Werner Kühlbrandt, Department of Structural Biology, Max Planck Institute for Biophysics
- Dr. Dirk Benndorf, Bioprocess Engineering, Max Planck Institute for Dynamics of Complex Technical Systems
- Prof. Dr. Rudolf Amann, Department of Molecular Ecology, Max Planck Institute for Marine Microbiology
- Dr. Ruben Garrido-Oter, Integrative Bioinformatics group, Max Planck Institute for Plant Breeding Research
- PD Dr. Heiko Liesegang, Institute of Microbiology and Genetics, Göttingen University
- Prof. Dr. Achim Kaasch, Institute for Medical Microbiology and Hospital Hygiene, Otto von Guericke University Magdeburg
- Prof. Dr. Lothar H. Wieler, Robert Koch Institute
- Prof. Dr. Franz Narberhaus, Microbial Biology, Ruhr-University Bochum
- Prof. Dr. Sören Auer; Dr. Irina Sens, Data Science & Digital Libraries Research Group, Technical Information Library (TIB)
- Prof. Dr. Jörn Stitz, Research Group Pharmaceutical Biotechnology, TH Köln -University of Applied Sciences
- Prof. Dr. Heinz Koepl, Centre for Synthetic Biology, Technical University of Darmstadt (TU Darmstadt)
- Dr. Ilias Lagkourdos, ZIEL - Institute for Food & Health, Technical University of Munich
- Prof. Dr. Maria J.G.T. Vehreschild, Clinical Microbiome Research Group, University Hospital Cologne
- Prof. Dr. Jochen Mattner, Institute of Microbiology – Clinical Microbiology, Immunology and Hygiene, University Hospital Erlangen
- Prof. Dr. Maria J.G.T. Vehreschild, Infectious Diseases, University Hospital Frankfurt
- Prof. Dr. Alexander Dilthey, Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne
- Prof. Dr. Alexander Dilthey, Institute of Medical Statistics and Computational Biology, University of Cologne
- Prof. Dr. Uwe Völker, Department of Functional Genomics, University of Greifswald
- Prof. Dr. Uwe T. Bornscheuer, Institute for Biochemistry, University of Greifswald
- Prof. Dr. Dörte Becher, Institute for Microbiology, University of Greifswald

- Prof. Dr. Wolfgang Streit, Microbiology & Biotechnology, University of Hamburg
- Prof. Dr. Martin Kaltenpoth, Max Planck Institute for Chemical Ecology
- Next Generation Sequencing Competence Network including
 - Prof. Dr. Joachim Schultze, Life & Medical Sciences Institute (LIMES) - I206 / The West German Genome Center (WGGC), University of Bonn
 - Dr. Marc Höppner, Institute of Clinical Molecular Biology (IKMB), Competence Centre for Genome Analysis Kiel (CCGA Kiel), University of Kiel
 - Prof. Olaf Riess, NGS Competence Center Tübingen (NCCT), University of Tübingen
 - Dr. Andreas Dahl, Dresden-Concept Genome Center (DcGC), University of Dresden

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

- 201-05 - Allgemeine Genetik und funktionelle Genomforschung
- 201-07 - Bioinformatik und Theoretische Biologie
- 202-03 - Organismische Interaktionen, chemische Ökologie und Mikrobiome pflanzlicher Systeme
- 203-03 - Ökologie und Biodiversität der Tiere und Ökosysteme, Organismische Interaktionen
- 204 - Mikrobiologie, Virologie und Immunologie
- 205-08 - Pharmazie
- 205-31 - Klinische Infektiologie und Tropenmedizin

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

NFID4Microbiota supports all fields of microbiology - including bacteriology, virology, protistology, mycology and parasitology - in the translation of research data into knowledge by developing and providing data management solutions, analytical tools, (meta-) data standards as well as training and support. The aim is to establish seamless and frictionless workflows from the data creation to the publication which will help researchers to focus on the generation of new insights to solve global problems. Following the idea of Open Science and the FAIR principles, NFID4Microbiota will establish solutions to make data of numerous types in the field of microbiology further manageable, findable as well as reusable and the research process as efficient, transparent and reproducible as possible. With 10 (co)-applicants, several participating institutions (including 3 DFG Priority Programmes, 3 DFG Collaborative Research Centers and one DFG Cluster of Excellence) from all fields of microbiology and the support of 5 national scholarly societies, NFID4Microbiota represents already now a large fraction of the German microbiology community. The interaction with numerous other NFDI consortia as well as with further national and international partners lays the foundation for efficient and sustainable development of solutions and the required cultural change in the community.

Microbiota (i.e. microbial communities), single microbial species (including bacteria, archaea, unicellular eukaryotes) and viruses have a strong influence on many aspects of human life, from individual health to climate change. In the face of wider global problems, our understanding of ecosystems needs to improve dramatically to be able to meet all the challenges posed by human activities. To date, microbes have been largely ignored. According to the UN, one of the greatest

medical threats is the increase in antibiotic resistance of bacterial pathogens, and the current COVID-19 pandemic shows the dramatic impact of viral infections on all areas of our lives. Both seemingly disconnected important issues require a better understanding of the world of microorganisms. In addition, countless species that have not yet been cultured are potential sources of substances with relevance for biotechnology and medicine that remain to be explored. One of the greatest challenges for the understanding of microbiota lies in the complexity of the numerous biotic interactions between specific organisms with any given microbiome and abiotic environmental factors. The study of individual species and entire communities, in particular the mapping and deciphering of molecular interactions with their functional portfolio and underlying regulatory mechanisms, is a crucial step towards understanding and more effective use of microbial species and prevention of diseases.

Research areas such as biomedicine, agriculture, marine research and biotechnology pursue very different scientific questions, but share the need for efficient management, storage, documentation and analysis of data on microbial organisms, microbiomes as well as their functional parts and molecular interactions between their members. While high-throughput methods can easily provide a wealth of omics data at different molecular levels (DNA, RNA, proteins, metabolites), e.g., to decipher the functioning of individual species or communities of microorganisms and their contributions to human health and environmental processes, the storage, analysis and integration of these data with available knowledge (e.g. persisting in literature) is not trivial. Often, researchers do not have the expertise and/or resources to deal with such data generated by high-throughput technologies. Furthermore, the increasing size and complexity of the data and the necessary processing, archiving, and integrative steps result in considerable demands.

The goal of the NFDI4Microbiota consortium is to make the storage, integration and analysis of data, especially omics data, for the study of microbial species and diverse microbiomes consistent, reproducible and accessible to all areas of life sciences. The consortium will support researchers with different scientific questions to understand individual microbial species as well as communities and the interaction between species within them and to make their own data accessible and reusable. NFDI4Microbiota will provide IT infrastructure, data management workflows, analytical tools, training and support for the community. Together with international partners and other NFDI consortia new standards will be developed. The consortium will thus enable efficient and reproducible processing of different types of data generated by high-throughput analytical devices, thereby helping to decipher microbial species and their interactions at the molecular level. The data will include genomes, transcriptomes, proteomes and metabolomes for individual species and the analogues for entire microbial communities: metagenomes, meta-transcriptomes, meta-proteomes and meta-metabolomic data. In addition, emerging analytical approaches for data from single cell sequencing and high-throughput imaging will be supported. Moreover, the consortium will make the data interoperable and enable the enrichment of these data with metadata from analytical instruments and databases as well as knowledge automatically extracted from the literature. In order to ensure a seamless data flow, NFDI4Microbiota will work with data generators such as sequencing facilities (the national Next Generation Sequencing Competence Network with its four centers is participant) and will promote direct storage of data and associated metadata in the infrastructure provided by the consortium. The data will be evaluated through analysis pipelines according to the wishes of the researchers who commissioned the analysis. Raw data, metadata, and the results will be stored in standardized form in customized repositories for long-term availability.

NFDI4Microbiota will fully comply with the FAIR principles (Findable, Accessible, Interoperable, Reusable) and promote Open Science with all its facets. Sensitive personal data will be treated with the necessary care. The consortium will define the necessary specifications for the rich collection of metadata describing the sample conditions and allowing a standardised and quality controlled submission of data. The members of our consortium are leading the development of the International Human Microbiome Standards (IHMS), and NFDI4Microbiota will work with the community to promote similar standards for other fields. In addition, the consortium will encourage participants to choose the most liberal licenses for the submitted data sets in order to avoid legal barriers to data sharing. Consistent management of the submitted data and rich annotation with metadata will be at the core of the powerful search for original data and results, as well as efficient reuse and comparison by the research community. To generate sustainable solutions, NFDI4Microbiota will provide software stacks based on FLOSS (Free/Libre/Open Source Software) following good software development practice. All developed software will be published under OSI (Open Source Initiative) compliant licenses.

The consortium will be in constant exchange with users to receive feedback and to adapt the services to the needs of its research community. As central hub we will facilitate the networking of the microbiology research community and its subfields and NFDI4Microbiota has already now support by 5 German scholarly societies from the fields of microbiology - DGfM, DGHM, DGP, GFV and VAAM. Furthermore, it will build bridges to related fields like agriculture, medicine and biotechnology. In addition, established links with international organizations such as ELIXIR, the European Bioinformatics Institute, GO FAIR Microbiome and The Carpentries as well as local partners like fdm.nrw will lay the foundation for an efficient implementation of solutions and numerous synergies.

NFDIMicrobiota bundles numerous measures in the following task areas:

- Community & Networking
- Standards & Policies
- Technical Infrastructure
- Services
- Use Cases
- Coordination & Communication

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

- The de.NBI Cloud compute and data storage infrastructure
- Numerous analytical software tools for the processing of data
- Numerous databases and terminology services
- Discovery services for literature and data
- Long-term preservation service
- Electronic Lab Notebook consulting and tools
- Data and information science/literacy training

Interfaces to other proposed NFDI consortia: brief description of existing agreements for collaboration and/or plans for future collaboration [max 1 page]

- **DataPLANT**: Collaboration on metadata and workflows for omics data and training.
- **GHGA**: Development of common standards and procedures for human microbiome data

- **NFDI4Agri**: Development of common metadata standards for sampling of microbial species/communities.
- **NFDI4BioDiversity**: Development of common metadata standards for sampling of microbial species/communities.
- **NFDI4Bioimage**: Collaboration on standards and workflows regarding imaging data from biological samples
- **NFDI4Chem**: Collaboration regarding Electronic Lab Notebooks
- **NFDI4DataScience**: Development on solution for Linked Open Data and Research Knowledge Graphs
- **NFDI4Health**: Standardisation of chemical components such as medication, dietary factor or metabolom data; Exchange regarding the inclusion of microbiome related data in medical studies.
- **NFDI4Immo**: Development of solution for the parallel analysis of microbiota and the immune system; Harmonization of metadata and ontologies describing the host as well as identifying and enhancing formalized descriptions of sampling procedures.
- **NFDI4Ing**: Collaboration regarding biotechnology related data
- **NFDI4Life Umbrella** and its sub-consortia: Collaborate on several cross-disciplinary issues of the life science community.
- **NFDI4RSE**: Collaboration on standards and recommendations regarding analysis workflows and training.

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

- Standards regarding data lineage/provenance
- Standards regarding high-throughput sequencing data
- Metadata standards for samples from medical and environmental sources
- Standard procedures for the handling of medical data with privacy issues/need for anonymisation
- Solutions for reproducible data analysis workflow
- Infrastructure for ordering, tracking and maybe billing of data generation/analyses
- Incentives for data publications
- Training of data and information science/literacy

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

For the above mentioned cross-topics we would mostly contribute by joining and/or facilitating the discussion and providing infrastructure for the implementation. We can also imagine that for certain software solution teams of software developers from different consortia a set up. For training the development of common open educational resources (OERs) for more general topics is planned which could also be done in cross-consortia teams.