

The National Research Data Infrastructure for the Research of Microbiota (NFDI4Microbiota) - Second Community Workshop

Anke Becker^{1,2}, Peer Bork^{1,3}, Boyke Bunk^{1,4}, Thomas Clavel^{1,5}, Ulisses Nunes da Rocha^{1,6}, Konrad U. Förstner^{1,7}, Alexander Goesmann^{1,8}, Manja Marz^{1,9}, Alice McHardy^{1,10}, Jörg Overmann^{1,4}, Alfred Pühler^{1,11}, Lorenz C. Reimer^{1,4}, Alexander Sczyrba^{1,11}, Jens Stoye^{1,11}, Thea Van Rossum^{1,3}, Justine Vandendorpe^{1,7}

¹ NFDI4Microbiota Consortium

² Philipps-Universität Marburg, Marburg, Germany

³ European Molecular Biology Laboratory, Heidelberg, Germany

⁴ Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany

⁵ RWTH Aachen University, Aachen, Germany

⁶ Helmholtz Centre for Environmental Research—UFZ, Leipzig, Germany

⁷ ZB MED - Information Centre for Life Sciences, Cologne, Germany

⁸ Justus-Liebig-University Giessen, Giessen, Germany

⁹ Friedrich Schiller University Jena, Jena, Germany

¹⁰ Helmholtz Centre for Infection Research, Brunswick, Germany

¹¹ Bielefeld University, Bielefeld, Germany

The German Council for Scientific Information Infrastructures (Rfll) and the Joint Science Conference (GWK) identified a lack of infrastructure to systematically manage scientific and research data in specific fields. To address this problem, the Rfll and the GWK initiated the National Research Data Infrastructure (NFDI), a German Federal and State Programme that will fund several consortia. The Rfll and the GWK define consortia as “*groups of users and providers of research data that come together for the purpose of long-term cooperation in a specific domain*”.

With this workshop we would like to present NFDI for the Research of Microbiota (NFDI4Microbiota, <https://nfdi4microbiota.de/>), share its vision and collect feedback from the community.

NFDI4Microbiota aims at making the analysis of multi-omics data related to microbial species and diverse microbiomes consistent, reproducible and accessible to all fields of the life science. The consortium will achieve this by assisting researchers in dealing with different scientific challenges to understand individual microbial species and communities, as well as interspecific interactions on a molecular level. NFDI4Microbiota will also provide the computational infrastructure, analytical tools and training for the research community to compile, analyse, store and enrich multi-omics data such as (meta-)genomes, (meta-)transcriptomes, (meta-)proteomes and (meta-)metabolomes. Researchers will have the opportunity to process their data through analysis workflows according to their wishes. Raw data, metadata and results will be deposited in repositories for long-term availability. Sensitive personal data will be treated with necessary care and will undergo pseudonymization. The consortium will fully comply with the FAIR (Findable, Accessible, Interoperable and Re-usable) principles, and will promote Open Science with all its facets. To achieve these goals NFDI4Microbiota will collaborate with existing infrastructure providers, researchers and other NFDI consortia to agree on standards and interfaces.

While a set of well-established institutions forms the core of NFDI4Microbiota, we are still looking for contributions from and collaborations with associated partners.

Keywords: high-throughput data analysis, multi-omics, bioinformatics pipeline, training, FAIR principles