Research Ideas and Outcomes 12: e215133
doi: 10.3897/rio.12.e215133

Grant Proposal

Second funding phase proposal: NFDI4Microbiota – National Research Data Infrastructure for Microbiota Research

Konrad U. Förstner‡, Sina-Victoria Barysch§, Anke Becker|, Michael Beckstette¶, Sebastian Beyvers#, Martin Bole¤, Peer Bork§, Lukas Brehm#, Noriko Cassman«, Thomas Clavel», Muhammad Elhossary‡, Frank Förster#, Alexander Goesmann#, Anandhi Iyappan§, Stefan M. Janssen#, Lina Junctorius‡, Sebastian Jünemann¶, Sanchita Kamath¤, Kassian Kobert¶, Julia Koblitz^, Marcus Lechner|, Maja Magel», Katharina Markus‡, Manja Marz«, Fernando Meyer˅, Philipp Münch˅, Ulisses Nunes Da Rocha¤, Carmen Paulmann˅, Daniel Podlesny§, Lorenz Reimer^, Till Sauerwein‡, Jannis Schlegel#, Isabel Schober^, Alexander Sczyrba¶, Jens Stoye¶, Clemens Thölken|, Justine Vandendorpe‡, Michael Vockenhuber|, Alice McHardy˅

‡ ZB MED - Information Centre for Life Sciences, Cologne, Germany
§ EMBL, Heidelberg, Germany
| Philipps-Universität Marburg, Marburg, Germany
¶ Bielefeld University, Bielefeld, Germany
# Justus-Liebig-University Gießen, Gießen, Germany
¤ Helmholtz Centre for Environmental Research (UFZ), Leipzig, Germany
« Friedrich-Schiller-University Jena, Jena, Germany
» RWTH University Hospital, Aachen, Germany
^ Leibniz Institute DSMZ German Collection of Microorganisms and Cell Cultures GmbH, Braunschweig, Germany
˅ Helmholtz-Centre for Infection Research (HZI), Braunschweig, Germany

Corresponding author: Konrad U. Förstner (foerstner@zbmed.de)

Reviewable v 1

Received: 10 Sep 2026 | Published: 05 Oct 2026

Citation: Förstner KU, Barysch S-V, Becker A, Beckstette M, Beyvers S, Bole M, Bork P, Brehm L, Cassman N, Clavel T, Elhossary M, Förster F, Goesmann A, Iyappan A, Janssen SM, Junctorius L, Jünemann S, Kamath S, Kobert K, Koblitz J, Lechner M, Magel M, Markus K, Marz M, Meyer F, Münch P, Nunes Da Rocha U, Paulmann C, Podlesny D, Reimer L, Sauerwein T, Schlegel J, Schober I, Sczyrba A, Stoye J, Thölken C, Vandendorpe J, Vockenhuber M, McHardy A (2026) Second funding phase proposal: NFDI4Microbiota – National Research Data Infrastructure for Microbiota Research. Research Ideas and Outcomes 12: e215133. https://doi.org/10.3897/rio.12.e215133

© Förstner K et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Modern microbiological research generates vast quantities of data, particularly from high-throughput "omics" technologies. These datasets promise unprecedented insights into microbial diversity, function, and interactions, yet they are often deposited in unstructured formats with inadequate metadata, significantly hampering their discoverability, interoperability, and long-term reusability. Adding to this challenge is a long tail of subdomain-specific data types that typically lack a consistent structure. NFDI4Microbiota empowers microbiology researchers by transforming their research data into actionable knowledge that offers deeper insights into microbial life. The consortium makes data, software tools, analytical workflows, and machine-learning models FAIR (Findable, Accessible, Interoperable, and Reusable), ensuring they are AI-ready and openly available for efficient reuse. Key strategies include comprehensive training, developing and disseminating community-driven standards and best practices, fostering strong data stewardship, providing user-friendly data management tools and infrastructure, and promoting standardisation across the microbiology community. Building on our successful first funding phase, we will expand NFDI4Microbiota's efforts in the next phase to enhance data accessibility and reuse while enabling AI applications. Our achievements and infrastructure from the first funding period provide a solid foundation for achieving an even greater impact across the microbiology research community and fostering a seamless flow of data. Community engagement remains vital: we will strengthen collaboration with existing stakeholders while actively reaching out to previously underrepresented groups, specifically researchers working on eukaryotic microbes such as fungi and protists, and we will work more closely with data-generating facilities to capture rich metadata early on. In addition, we will further strengthen our portfolio of services, databases, and analytical tools, ensuring that they are robust, interoperable, and tailored to evolving community needs. These enhancements will support a more comprehensive and inclusive data infrastructure for microbiological research. To help realise the One NFDI and EOSC vision of a coherent, sustainable, and collaborative research data ecosystem across all disciplines, NFDI4Microbiota will actively engage in cross-consortial initiatives, foster interoperability, and apply the unique perspectives of microbiological science to help shape the broader NFDI landscape.

Keywords

Research data management, FAIR principles, microbiota, NFDI, Germany

1 General Information

Table 1 lists abbreviations used throughout the text.

Table 1.
List of all abbreviations.

(T)RE	(Trusted) Research Environments
AAI	Authentication and Authorization Infrastructure
AI	Artificial Intelligence
AMR	Antimicrobial Resistance
API	Application Programming Interface
AU	Administration Unit
Base4NFDI	Basic services for NFDI
BGC	Biosynthetic Gene Cluster
BIBI	Bielefeld Institute for Bioinformatics Infrastructure
BIC	Bioinformatics Core Facility
BoD	Board of Directors
CA	Consortium Agreement
CAMI	Critical Assessment of Metagenome Interpretation
CBD	Convention on Biological Diversity
CIG	Community Interest Group
CLI	Command Line Interface
CLoWM	Cloud-based Workflow Manager
CLUE-TERRA	CoLaborative Multi-domain Exploration of TERRestriAl metagenomes
CoC	Code of Conduct
COSI	"Microbiome" Community of Special Interest
CPB	Centre for Pathogen Bioinformatics
CPU	Central Processing Unit
CRC	Collaborative Research Centres
DALIA4NFDI	Data and Language Infrastructure for NFDI
DAPHNE4NFDI	Data from PHoton and Neutron Experiments for NFDI
DataPLANT	Data in PLANT research
de.KCD	German Competence Center Cloud Technologies for Data Management and Processing
de.NBI	German Network for Bioinformatics Infrastructure
DFG	German Research Foundation

DGfM	German Society for Mycology
DGHM	German Society for Hygiene and Microbiology
DGP	German Society for Parasitology
DKZ.2R	Data Literacy Center Rhine-Ruhr
DMP	Data Management Plan (DMP)
DMP4NFDI	Data Management Plan for NFDI
DMykG	German-Speaking Mycological Society
DNA	Deoxyribonucleic Acid
DOI	Digital Object Identifier
DOID	Disease Ontology ID
DRS	Data Repository Service
DSI	Digitial Sequence Information
DSMZ	German Collection of Microorganisms and Cell Cultures
DZIF	German Center for Infection Research
EBI	European Bioinformatics Institute
ECCB	European Conference on Computational Biology
EDIS	Equality, Diversity, and Inclusion in Science and Health
EGA	European Genome-phenome Archive
EICAT	EMBL's International Centre for Advanced Training
ELIXIR	European life-sciences infrastructure for biological information
ELIXIR-DE	ELIXIR's German Node
ELLS	European Learning Laboratory for the Life Sciences
ELN	Electronic Lab Notebook
ELSA	Ethical, Legal, and Social Aspects
EMBL	European Molecular Biology Laboratory
ENA	European Nucleotide Archive
EOSC	European Open Science Cloud
ESFRI	European Strategy Forum on Research Infrastructures
EuMik	Society for Eukaryotic Microbiology
EVBC	European Virus Bioinformatics Centre
FAIR	Findable, Accessible, Interoperable, and Re-usable

FAIRagro	FAIR Data Infrastructure for Agrosystems
FAQ	Frequently Asked Question
FDM.NRW	Landesinitiative Forschungsdatenmanagement Nordrhein-Westfalen
FDO	FAIR Digital Object
Flex Fund	Flexible Fund
FLI	Leibniz Institute on Aging – Fritz Lipmann Institute
FLOSS	Free Libre Open Source Software
FRL	PUBLISSO – Repository for Life Sciences
FSU Jena	Friedrich Schiller University Jena
FTE	Full-Time Equivalent
GA	General Assembly
GA4GH	Global Alliance for Genomics & Health
GDPR	General Data Protection Regulation
GfV	German Society for Virology
GHGA	German Human Genome-Phenome Archive
GPS	Global Positioning System
GPU	Graphics Processing Unit
GSC	Genomic Standards Consortium
GTDB	Genome Taxonomy Database
GUI	Graphical User Interface
HeFDI	Hessian Research Data Infrastructure
HPC	High-performance computing
HPO	Human Phenotype Ontology
HZI	Helmholtz-Zentrum für Infektionsforschung
IAM4NFDI	Identity & Access Management for the German National Research Data Infrastructure
IHGC	International Human Genome Consortium
INSDC	International Nucleotide Sequence Database Collaboration
ISCB	International Society for Computational Biology
ISMB	Intelligent Systems for Molecular Biology
IT	Information Technology
IZKF	Interdisciplinary Center for Clinical Research

JLU	Justus Liebig University Giessen
K8s	Kubernetes
KGI4NFDI	Knowledge Graph Infrastructure for the German National Research Data Infrastructure
KPI	Key Performance IndicatorKey Performance Indicator
LIMS	Laboratory information management systems
LIV	Leibniz Institute of Virology
LLM	Large Language Model
LTA	Long-term access and preservation
M4C	Microbes-for-Climate
MAG	Metagenome-Assembled Genome
MaRC3a	Marburg Computing Cluster
MaRDI	Mathematics Research Data Initiative
MaSC	Marburg Storage Cluster
MCP	Model Context Protocols
MIxS	Minimum Information about any (x) Sequence
ML	machine learning
MPIterMIc	Max Planck Institute for Terrestrial Microbiology
MSI	Metabolomics Standards Initiative
NAKO	National Cohort Study
NCBI	National Center for Biotechnology Information
NFDI	National Research Data Infrastructure
NFDI-MatWerk	National Research Data Infrastructure for Materials Science and Materials Engineering
NFDI4Biodiversity	Research Data Infrastructure for Biodiversity, Ecology and Environmental Data
NFDI4BIOIMAGE	National Research Data Infrastructure for Bioimaging Data (Microscopy and Image Analysis)
NFDI4Cat	National Research Data Infrastructure for Catalysis-Related Sciences
NFDI4Chem	National Research Data Infrastructure for Chemistry
NFDI4Culture	National Research Data Infrastructure for Research Data on Material and Immaterial Cultural Heritage
NFDI4DataScience	National Research Data Infrastructure for Data Science and Artificial Intelligence
NFDI4Earth	National Research Data Infrastructure for Earth System Sciences
NFDI4Health	National Research Data Infrastructure for Personal Health Data
NFDI4Immuno	National Research Data Infrastructure for Immunology

NFDI4Ing	National Research Data Infrastructure for Engineering Sciences
NFDI4Memory	National Research Data Infrastructure for the Historically Oriented Humanities
NFDI4Microbiota	National Research Data Infrastructure for Microbiota Research
NFDI4Objects	National Research Data Infrastructure for the Material Remains of Human History
NGS	Next Generation Sequencing
NGS-CN	Next Generation Sequencing Competence Network
NMDC	National Microbiome Data Collaborative
NRW	North Rhine-Westphalia
OBO	Open Biological and Biomedical Ontologies
OER	Open Educational Resource
ORCID	Open Researcher and Contributor ID
PID	Persistent Identifier
PID4NFDI	PID Infrastructure for NFDI
PR	Public Relation
PRB	Project Review Board
PSI	Proteomics Standards Initiative
PUNCH4NFDI	Particles, Universe, NuClei and Hadrons for the National Research Data Infrastructure
RAG	Retrieval-Augmented Generation
RAM	Random Access Memory
RDA	Research Data Alliance
RDC	Research Data Commons
RDF	Resource Description Framework
RDM	Research Data Management
RDMO	Research Data Management Organizer
RDMTraining4NFDI	Research Data Management Training for the National Research Data Infrastructure
REST	Representational State Transfer
RO-Crate	Research Object Crate
ROR	Research Organization Registry
RWTH	Rheinisch-Westfälische Technische Hochschule Aachen
SAB	Scientific Advisory Board
SEV	Secure Encrypted Virtualization

SFB	Sonderforschungsbereich
SME	Secure Memory Encryption
SOP	Standard Operating Procedure
SPE	Secure Processing Environments
SPP	Priority Programmes
SRA	Sequence Read Archive
TA	Task Area
TES	Task Execution Service
TESS	Training eSupport System
Text+	Language and text-based research data infrastructure
TREC	Traversing European Coastlines
TS4NFDI	Terminology Services for the German National Research Data Infrastructure
UFZ	Helmholtz Centre for Environmental Research
UKA	University Hospital of RWTH Aachen
UMR	University of Marburg
UN	United Nations
UNIBI	Bielefeld University
USIG	User Special Interest Group
VAAM	Association for General and Applied Microbiology
VAT	Visualization, Analysis and Transformation Tool
ViBiom	Virus Bioinformatics meeting
WHO	World Health Organisation
WMS	Workflow Management System
WR	German Council of Science and Humanities
ZB MED	ZB MED - Information Centre for Life Sciences
zedif	Competence Center Digital Research

1.1 Name of the consortium in English and German

NFDI4Microbiota - National Research Data Infrastructure for Microbiota Research

NFDI4Microbiota - Nationale Forschungsdateninfrastruktur für Mikrobiota-Forschung

1.2 Summary of the proposal in German

Die moderne mikrobiologische Forschung erzeugt enorme Mengen an Daten, insbesondere durch Hochdurchsatz-Technologien ("Omics"). Diese Datensätze versprechen tiefgreifende Einblicke in mikrobielle Diversität, Funktionen und Interaktionen. Häufig werden sie jedoch in unstrukturierter Form mit unzureichenden Metadaten abgelegt, was ihre Auffindbarkeit, Interoperabilität und langfristige Nachnutzbarkeit erheblich einschränkt. Hinzu kommt eine Vielzahl fachspezifischer Datentypen mit uneinheitlicher Struktur. NFDI4Microbiota möchte Forschende dabei unterstützen, ihre Forschungsdaten in verwertbares Wissen zu überführen und so ein tieferes Verständnis mikrobiellen Lebens zu ermöglichen. Dazu macht das Konsortium Daten, Softwaretools, Analyse-Workflows und maschinelle Lernmodelle zu FAIR (findable, accessible, interoperable, reusable) Ressourcen – KI-tauglich und offen nachnutzbar. Zentrale Strategien sind Schulungsangebote, die Entwicklung und Verbreitung gemeinschaftsgetragener Standards und Best Practices, die Förderung einer Kultur des Data Stewardship, benutzerfreundliche Werkzeuge und Infrastrukturen für das Datenmanagement sowie die Etablierung standardisierter Verfahren in der mikrobiologischen Forschung. Aufbauend auf den Erfolgen der ersten Förderphase intensivieren wir unsere Aktivitäten, um Datenzugänglichkeit und -nachnutzung weiter zu verbessern und uns als KIEnabler zu positionieren. Die in der ersten Förderphase erreichten Meilensteine und etablierten Infrastrukturen bilden eine solide Basis, um unsere Wirkung in der Forschungsgemeinschaft auszuweiten und einen reibungslosen Datenfluss zu fördern. Die Einbindung der Community bleibt zentral: Wir stärken die Zusammenarbeit mit bestehenden Partnern und binden gezielt bislang unterrepräsentierte Gruppen ein, insbesondere Forschende zu eukaryotischen Mikroben wie Pilzen und Protisten. Zudem arbeiten wir enger mit datenproduzierenden Einrichtungen zusammen, um bereits früh hochwertige Metadaten zu erfassen. Unser Portfolio an Diensten, Datenbanken und Analysewerkzeugen wird weiter ausgebaut, um robust, interoperabel und an den sich wandelnden Bedarf der Community angepasst zu sein. So entsteht eine umfassendere und inklusivere Dateninfrastruktur für die mikrobiologische Forschung. Zur Verwirklichung der One-NFDI- und EOSC-Vision eines disziplinübergreifenden, nachhaltigen und kollaborativen Forschungsdatenökosystems engagiert sich NFDI4Microbiota aktiv in konsortienübergreifenden Initiativen, fördert Interoperabilität und bringt mikrobiologische Perspektiven in die Weiterentwicklung des NFDI-Gesamtverbunds ein.

1.3 Applicant institution

Table 2 lists the applicant institution.

Table 2.
Applicant institution and location.

Applicant institution	Location
ZB MED – Information Centre for Life Sciences	Cologne

1.4 Spokesperson

Table 3 lists the spokesperson.

Table 3.
Spokesperson and institutional affiliation.

Spokesperson	Institution, location
Prof. Dr. Konrad Förstner	ZB MED, Gleueler Straße 60, 50931, Cologne

1.5 Co-applicant institutions

Table 4 lists the co-applicant institutions.

Table 4.
Co-applicant institutions and their locations.

Co-applicant institutions	Location
Bielefeld Institute for Bioinformatics Infrastructure, Bielefeld University(UNIBI)	Universitätsstraße 25,33615 Bielefeld
European Molecular Biology Laboratory (EMBL)	Meyerhofstraße 1,69117 Heidelberg
Friedrich Schiller University Jena (FSU Jena)	Fürstengraben 1,07743 Jena
Helmholtz Centre for Environmental Research (UFZ)	Permoserstraße 15,04318 Leipzig
Helmholtz Centre for Infection Research (HZI)	Inhoffenstraße 7,38124 Braunschweig
Justus-Liebig-University Gießen (JLU Gießen)	Ludwigstraße 23,35390 Gießen
Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures	Inhoffenstraße 7B,38124 Braunschweig
Philipps-Universität Marburg (UMR)	Biegenstraße 10,35037 Marburg
RWTH Aachen University (RWTH)	Templergraben 55,52062 Aachen

1.6 Co-spokespersons

Table 5 lists the co-spokespersons.

Table 5.
Co-spokespersons, their affiliated institutions, and task area(s).

Co-spokespersons	Institution, location	Task area(s)
Prof. Dr. Alexander Sczyrba & Prof. Dr. Jens Stoye	Bielefeld University, Bielefeld	TA1, TA2, TA3, TA4
Prof. Dr. Peer Bork	EMBL, Heidelberg	TA1, TA2, TA3, TA4
Prof. Dr. Manja Marz	FSU Jena, Jena	TA1, TA2, TA3
Dr. Ulisses Nunes da Rocha	UFZ, Leipzig	TA1, TA2, TA3, TA4
Prof. Dr. Alice C. McHardy	HZI, Braunschweig	TA1, TA2, TA3, TA4, TA5
Prof. Dr. Alexander Goesmann	JLU Gießen, Gießen	TA1, TA2, TA3, TA4, TA5
Dr. Lorenz Reimer	DSMZ, Braunschweig	TA1, TA2, TA3, TA4
Prof. Dr. Anke Becker	UMR, Marburg	TA1, TA2, TA3, TA4, TA5
Prof. Dr. Thomas Clavel	RWTH Aachen, Aachen	TA1, TA2, TA3

1.7 Participants

Table 6 lists the participating institutions.

Table 6.
Participating Institutions

Participating Institution	Location	ROR	Contribution
Anhalt University of Applied Sciences	Köthen	0076zct58	Expertise in improving metadata standards for metaproteomics and bioinformatic analysis of metaproteome data; standardisation of experimental protocols; provision of training; nomination of an ambassador
Bergische Universität Wuppertal	Wuppertal	00613ak93	Expertise in viral metadata collection; standardisation of experimental protocols; provision of training; nomination of an ambassador
de.NBI e.V.	Heidelberg	-	Expertise in bioinformatics infrastructure and training; nomination of an ambassador

Participating Institution	Location	ROR	Contribution
Federal Institute for Materials Research and Testing (BAM)	Berlin	03x516a66	Expertise in material-microbiome interactions; provision of microbiota datasets for surveillance of biofilms; connections to Joint Genome Institute (JGI) and other NFDI consortia; nomination of an ambassador
Freie Universität Berlin	Berlin	046ak2485	Expertise in secondary metabolomics data and their genomic and transcriptomic basis; evaluation of metabolomics data; nomination of an ambassador
German National Library of Science and Technology (TIB)	Hannover	04aj4c181	Provision of service infrastructure such as the Open Research Knowledge Graph; expertise in semantic data modelling, persistent identifiers, long-term archiving and knowledge graph development; provision of training; nomination of an ambassador
GFBio - German Federation for Biological Data e.V.	Bremen	02qtnz251	Cross-community coordination through NFDI4Biodiversity, Biodata and the Geo-Chem-Life Science Helpdesk; development and promotion of services; systematic information exchange; nomination of an ambassador
Goethe University Frankfurt	Frankfurt am Main	04cvxnb49	Expertise in viral metadata curation; standardisation of experimental procedures; provision of training; nomination of an ambassador
Hannover Medical School	Hannover	00f2yqf98	Expertise in multi-resistant bacteria sequencing and intestinal microbiomes
Helmholtz Centre Potsdam - GFZ German Research Centre for Geosciences	Potsdam	04z8jg394	Expertise in prokaryotic metadata collection; development of analysis tools; provision of training events
Hessian State Office for Health and Care (HLfGP)	Darmstadt	-	Expertise in bacterial and metagenome metadata collection; expertise as a public health organisation; standardisation of experimental protocols; provision of training; nomination of an ambassador
Jülich Research Centre	Jülich	02nv7yv05	Expertise in computational genomics and metagenomics; provision of compute and storage resources through de.NBI Cloud; provision of training
Julius Maximilian University Würzburg	Würzburg	00fbnyb24	Expertise in fungal and bacterial metadata collection; standardisation of experimental protocols; provision of training events
Karlsruhe Institute of Technology	Karlsruhe	04t3en479	Expertise in prokaryotic single-cell omics technologies; standardisation of experimental protocols and training; nomination of an ambassador

Participating Institution	Location	ROR	Contribution
Kiel University	Kiel	04v76ef78	Expertise in microbiota-focused multi-omics data integration; development of experimental protocols; provision of training; nomination of an ambassador
Leibniz Institute for Analytical Sciences - ISAS	Dortmund	02jhqqg57	Expertise in metaproteomics, research data management for metaproteomics and knowledge graph integration; nomination of an ambassador
Leibniz Institute for Natural Product Research and Infection Biology (HKI)	Jena	055s37c97	Expertise in microbiome-based research in the context of fungal infections; nomination of an ambassador
Leibniz Institute for New Materials (INM)	Saarbrücken	00g656d67	Standardisation of metadata schemas for Engineered Living Materials (ELMs); nomination of an ambassador
Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB)	Berlin	01nftxb06	Expertise in aquatic microorganisms; initiation of international collaborations funded by BMBF, DFG and EU; nomination of an ambassador
Leibniz Institute on Aging - Fritz Lipmann Institute (FLI)	Jena	039a53269	Expertise in microbiome research in the context of aging; use of NFDI4Microbiota resources for heterogeneous microbiome data management across research groups
Ludwig Maximilian University of Munich	Munich	05591te55	Expertise in the analysis of multi-omics approaches in parasitic protozoa; nomination of an ambassador
Max Delbrück Center (MDC)	Berlin	04p5ggc03	Expertise in microbiome analyses with Charité – Universitätsmedizin Berlin; development of multi-omics data integration software; collaboration with national and international networks; contribution to the GUT, DATA-SCI, CLINMIC, PATHO and MULTI use cases; nomination of an ambassador
Max-Planck-Gesellschaft zur Förderung der Wissenschaften e.V., Max Planck Institute for Evolutionary Biology	Plön	0534re684	Development of OpenBIS (ELN and LIMS system); expertise in microscopy and bioimage research data management; development of a knowledge graph integrating in-house and public data resources; connections to NFDI4BIOIMAGE; nomination of an ambassador
Max-Planck-Gesellschaft zur Förderung der Wissenschaften e.V., Max Planck Institute for Plant Breeding Research	Cologne	044g3zk14	Expertise in microbial synthetic communities, experimental standardisation and plant microbiota research; provision of training; nomination of an ambassador

Participating Institution	Location	ROR	Contribution
Robert Koch Institute (RKI)	Berlin	01k5qnb77	Expertise in bacterial pathogens and microbiome research; connections to national and international partners; contribution of proteomics capabilities
Technical University of Munich	Munich	02kkvpp62	Expertise at the interface between environmental and human microbiota; standardisation of microbiome analysis pipelines; application of recent developments in bioinformatics; nomination of an ambassador
TH Köln - University of Applied Sciences	Cologne	014nnvj65	Promotion of tools and services; provision of workshops
University of Bonn	Bonn	041nas322	Expertise in molecular mechanisms and functions of microbial biochemistry and cell biology
University of Cologne	Cologne	00rcxh774	Expertise in whole genome sequencing, antimicrobial resistance analysis, bacterial typing, plasmids and mobile genetic elements in infectious diseases, and interpretation of resistance determinants; standardisation of experimental protocols; provision of training; nomination of an ambassador
University of Duisburg-Essen	Duisburg-Essen	04mz5ra38	Expertise in research data management and training; nomination of an ambassador
University of Erlangen-Nuremberg (FAU)	Erlangen	00f7hpc57	Expertise in microbiome metadata collection; standardisation of experimental protocols; nomination of an ambassador
University of Hamburg	Hamburg	00g30e956	Expertise in microbial community analysis of environmental and algal microbiomes; provision of standardised experimental protocols and training
University of Hohenheim	Stuttgart	00b1c9541	Expertise in the definition and standardisation of methods and protocols for strain definition, identification and comparison; provision of training
University of Potsdam	Potsdam	03bnmw459	Expertise in metabolomic metadata collection; provision and standardisation of experimental protocols; nomination of an ambassador
University of Regensburg	Regensburg	01eezs655	Expertise in archaeal research; contribution of omics datasets; participation in feedback loops; nomination of an ambassador
University of Tübingen	Tübingen	03a1kwz48	Development of best-practice pipelines; expertise in metadata models for omics and imaging data; provision of training; nomination of an ambassador

Table 7 lists the participating networks.

Table 7.
Participating Networks

Network	Speaker	Location	GEPRIS	Contribution
CRC 1182: Origin and Function of Metaorganisms	Prof. Dr. Hinrich Schulenburg	Kiel	261376515	Expertise in integrative host-microbe research across multiple biological systems; collaboration among researchers; nomination of an ambassador
CRC 1371: Microbiome Signatures - Functional Relevance in the Digestive Tract	Prof. Dr. Dirk Haller	Munich	395357507	Expertise in building data analysis pipelines; nomination of an ambassador
CRC 1382: Gut-Liver Axis - Functional Circuits and Therapeutical Targets	Prof. Dr. Oliver Pabst	Aachen	403224013	Contribution to research data management plans; metadata harmonisation; multi-omics data integration; improvement of electronic lab notebooks; development of AI-based workflows; nomination of dedicated staff members
CRC 1583: Decisions in Infectious Diseases (DECIDE)	Prof. Dr. Thomas Rudel	Würzburg	492620490	Expertise in fungal and bacterial metadata collection; standardisation of experimental protocols; provision of training; nomination of an ambassador
CRC 2330: New concepts in prokaryotic virus-host interactions - from single cells to microbial communities	Prof. Dr. Julia Frunzke	Jülich	441953848	Expertise in bacterial and archaeal viruses; national and international collaborations; nomination of an ambassador
EXC 2051: Balance of the Microverse	Prof. Dr. Kirsten Küsel	Jena	390713860	Expertise in microbial ecology, data generation and data integration; contribution of datasets
EXC 2124: Controlling Microbes to Fight Infections (CMFI)	Prof. Dr. Andreas Peschel	Tübingen	390838134	Development of pipelines, platforms, and tools, including clone-specific microbiome analysis
EXC 2155: RESIST - Resolving Infection Susceptibility	Prof. Dr. Reinhold Förster	Hannover	390874280	Expertise in the molecular basis of immune deficiencies and the infectious agents involved
Fraunhofer Cluster of Excellence Immune-Mediated Diseases (CIMD)	Prof. Dr. Dr. Gerd Geisslinger	Frankfurt am Main	-	Expertise in preclinical and clinical research; initiation of national and international collaboration

Network	Speaker	Location	GEPRIS	Contribution
Global Microbiome Conservancy	Prof. Dr. Mathieu Groussin	Kiel	-	Expertise in managing worldwide microbiome data in compliance with international regulations on data usage and sharing; nomination of an ambassador
Microbial Community Database (MiCoDa)	Dr. Stephanie Jurburg	Leipzig	-	Expertise as the world's largest curated 16S rRNA gene database; provision of global outreach and training; nomination of an ambassador
Research Center One Health Ruhr	Prof. Dr. Alexander J. Probst	Duisburg-Essen	-	Expertise in environmental genomics and virusFISH; provision of standardised protocols and training; nomination of an ambassador
SPP 2474: Illuminating gene functions of the human gut microbiome	Prof. Dr. Lisa Maier	Tübingen	540841827	Expertise in abundant and functionally important bacterial members of the healthy human gut microbiome; nomination of an ambassador

Table 8 lists the participating individuals.

Table 8.
Participating Individuals

Participating Individual	Affiliated Institution	Location	ORCID	Contribution
Dr. Magdalene Cyra-Wolf	University of Duisburg-Essen, Landesgeschäftsstelle für Forschungsdatenmanagement (fdm.nrw)	Duisburg-Essen	0000-0001-7738-2703	Provision of communication infrastructure and RDM workshops as a central hub of research data management in North Rhine-Westphalia
Prof. Dr. Thomas Dandekar	Julius Maximilian University Würzburg, Chair of Bioinformatics, 00fbnyb24	Würzburg	0000-0003-1886-7625	Expertise in fungal and bacterial metadata collection; standardisation of experimental protocols; provision of training in the context of FungiNet (CRC 124) and DECIDE (CRC 1583); nomination of an ambassador

Participating Individual	Affiliated Institution	Location	ORCID	Contribution
Prof. Dr. Stephanie Deinhardt-Emmer	Jena University Hospital, Institute for Medical Microbiology, 035rzkx15	Jena	0000-0003-4495-4052	Provision of workflows and infrastructure best practices; clinical Use Case development and pilot studies; implementation of NFDI standards in hospital-based research
Apl. Prof. Dr. Elena Evguenieva-Hackenberg	Justus-Liebig-Universität Giessen (JLU Giessen), Institute for Microbiology and Molecular Biology, 033eqas34	Gießen	0000-0001-7270-3168	Expertise in model soil bacteria; standardisation of experimental protocols; nomination of an ambassador
PD Dr. Dimitrios Frangoulidis	Bundeswehr Institute of Microbiology, 01xexwj76	Munich	0000-0002-2864-3820	Contribution of database content; creation of information and training materials
Prof. Dr. W. Florian Fricke	University of Hohenheim, Department of Microbiome & Applied Bioinformatics, 00b1c9541	Stuttgart	0000-0003-1586-0962	Expertise in the definition and standardisation of methods and protocols for strain definition, identification and comparison; provision of training; nomination of an ambassador
Prof. Dr. Marco Galardini	Center for Experimental and Clinical Infection Research (TWINCORE), Microbial Pangenomes Lab, 04bya8j72	Hannover	0000-0003-2018-8242	Expertise in statistical methods applied to collections of annotated bacterial genomes; provision of training; nomination of an ambassador
Dr. Martin Gand	Justus-Liebig-Universität Giessen (JLU Giessen), Institute of Food Chemistry and Food Biotechnology, 033eqas34	Gießen	0000-0001-8211-691X	Expertise in fungal isolates and multi-omics analysis; sharing of fungal multi-omics data; nomination of an ambassador
PD Dr. Torsten Hain	Justus-Liebig-University Giessen (JLU Giessen), Institute of Medical Microbiology, 033eqas34	Gießen	0000-0003-0149-3137	Expertise in bacterial metadata collection; standardisation of experimental protocols; provision of training; nomination of an ambassador
Dr. Nico Jehmlich	Helmholtz Centre for Environmental Research - UFZ, Microbiome Biology, 000h6jb29	Leipzig	0000-0002-5638-6868	Expertise in environmental microbiomes and multi-omics approaches; nomination of an ambassador

Participating Individual	Affiliated Institution	Location	ORCID	Contribution
Dr. Frank Löffler	Friedrich Schiller University Jena (FSU Jena), Competence Center Digital Research (zedif), 05qpz1x62	Jena	0000-0001-6643-6323	Provision of training and education; development of reusable workflows; enhancement of local support structures
Prof. Dr. Kai Papenfort	Friedrich Schiller University Jena (FSU Jena), Chair of General Microbiology, 05qpz1x62	Jena	0000-0002-5560-9804	Expertise in planning and implementing high-throughput RNA sequencing technologies and analysis tools; contribution to the MULTI Use Case; nomination of an ambassador
Prof. Dr. Jana Seifert	University of Hohenheim, Department of Functional Microbiology of Livestock, 00b1c9541	Stuttgart	0000-0002-7690-8539	Expertise in livestock microbiome data collection; standardisation of species-specific analysis procedures for all omics approaches
Prof. Dr. Cynthia Sharma	Julius Maximilian University Würzburg, Research Center for Infectious Diseases (ZINF) and Department of Molecular Infection Biology II, 00fbnyb24	Würzburg	0000-0002-2321-9705	Expertise in diverse human pathogens and deep sequencing technologies; interaction with the NUCLEATE Cluster of Excellence; contribution to CRISPR and MULTI Use Cases; provision of training; nomination of an ambassador
Prof. Dr. Jörg Vogel	Helmholtz Institute for RNA-based Infection Research (HIRI), 02a98s891	Würzburg	0000-0003-2220-1404	Expertise in RNA-based approaches to prevent diseases; contribution to DATA-SCI and MULTI Use Cases; nomination of an ambassador
Prof. Dr. John Ziebuhr	Justus-Liebig-Universität Giessen (JLU Giessen), Institute of Medical Virology, 033eqas34	Gießen	0000-0002-5741-8825	Expertise in viral metadata collection; standardisation of experimental protocols; provision of training; nomination of an ambassador

1.8 Names and numbers of the DFG review boards (DFG-Fachkollegien) that reflect the subject orientation of the proposed consortium

2.11-05 General Genetics and Functional Genome Biology, 2.11-07 Bioinformatics and Theoretical Biology, 2.12-01 Evolution and Systematics of Plants and Fungi, 2.12-02 Ecology and Biodiversity of Plants and Ecosystems, 2.12-03 Organismic Interactions, Chemical Ecology and Microbiomes of Plant Systems, 2.13-03 Ecology and Biodiversity of Animals and Ecosystems, Organismic Interactions, 2.21 Microbiology, Virology and Immunology, 2.22-01 Epidemiology and Medical Biometry/Statistics, 2.22-02 Public Health, Healthcare Research, Social and Occupational Medicine, 2.22-05 Nutritional Sciences, 2.22-07 Medical Informatics and Medical Bioinformatics, 2.22-09 Pharmacology, 2.22-15 Gastroenterology, 2.22-22 Clinical Immunology and Allergology, 2.22-31 Clinical Infectiology and Tropical Medicine, 2.31-01 Soil Sciences, 3.46-01 Hydrogeology, Hydrology, Limnology, Urban Water Management, Water Chemistry, Integrated Water Resources Management, 4.21-04 Biological Process Engineering, 4.31-03 Coating and Surface Technology, 4.31-06 Polymeric and Biogenic Materials and Derived Composites, 4.41-06 Biomedical Systems Technology, 4.43-02 Software Engineering and Programming Languages, 4.43-06 Data Management, Data-Intensive Systems, Computer Science Methods in Business Informatics.

2 Scope and Objectives

2.1 Research domains or research methods addressed by the consortium, specific aim(s)

Microbial species—including bacteria, archaea, unicellular eukaryotes and viruses—influence every aspect of human life, from health and climate to agriculture and biotechnology. Microbiology is a multidisciplinary field that involves the study of microorganisms and their interactions with each other and the environment. Microbial research spans a correspondingly broad range of topics, ranging from basic molecular methodologies to multi-omics and functional analyses, and it generates vast amounts of data worldwide, especially through the application of high-throughput molecular technologies known as "omics" (Vailati-Riboni et al. 2017). Unfortunately, the resulting datasets and workflows often lack crucial metadata, making them hard to understand, reproduce, or reuse. Moreover, much of the information is stored in unstructured formats, further diminishing its value. The National Research Data Infrastructure for Microbiota Research (NFDI4Microbiota) helps microbiologists turn research data into comprehensive insights by connecting them with Findable, Accessible, Interoperable, and Re-usable (FAIR) (Wilkinson et al. 2016) and open solutions for managing data and workflows. The consortium offers tools and methods for mobilising, structuring, and linking available data to create a seamless flow throughout the research process. By making data, software, analytical workflows, and machine-learning models FAIR and open, we ensure they are ready for new approaches based on Artificial Intelligence (AI) (AI-readiness). Our target groups include not only traditional microbiologists but also

scientists from numerous connected fields, such as clinicians and biotechnologists. During the first funding period (see Fig. 1 and the initial proposal, Förstner et al. 2023), NFDI4Microbiota created numerous domain-specific services for the community (4.4 Services provided by the consortium), such as StrainInfo, which resolves strain identifiers in culture collections. It also developed solutions that other consortia can build on to create their own services, such as the domain-agnostic data storage and orchestration system Aruna (M4.2 - Data Storage Platform) and the Cloud based Workflow Manager (CloWM) (M4.3 - Workflow Execution Platform). Our training programme addressed a broad range of topics and has upskilled more than 2300 participants from diverse backgrounds (M1.1 - Training & Education). In the second funding period, we will build upon these achievements and expand our activities. As well as continuing our training programme and further developing existing services, we will extend new services and outreach activities to subcommunities that have received less attention so far, following the recommendation b) for the second funding period by the NFDI Expert Committee (NFDI Expert Committee 2023). While bacteriologists, virologists and microbiome researchers are already well served by our activities, we will now broaden our scope to include support for researchers working on eukaryotic microbial species. To this end, we will collaborate closely with specific communities and establish new resources such as the MicrobeDive database.

Figure 1.

Selection of NFDI4Microbiota's achievements and offerings in the first funding period in numbers; community networking activities include presentations, social media interactions, hackathons, and training events

2.2 Objectives and measuring success

NFDI4Microbiota's overarching goal in the second funding phase remains the same as in the first one: to drive the digital transformation of microbiology research by removing friction and accelerating the discovery of significant new insights. Achieving this goal is far from trivial, as the microbiological community is highly heterogeneous in its aims, data types, and practices. Nevertheless, the potential benefits are substantial, as microbiology is deeply interconnected with various other disciplines. During the second funding phase NFDI4Microbiota will therefore pursue the following key objectives (see also Fig. 2) to improve the Research Data Management (RDM) ecosystem for microbiology and to make data, software, models, and other output FAIR, open, and AI-ready. We recognise that these are long-term goals that will not be fully achieved within this period.

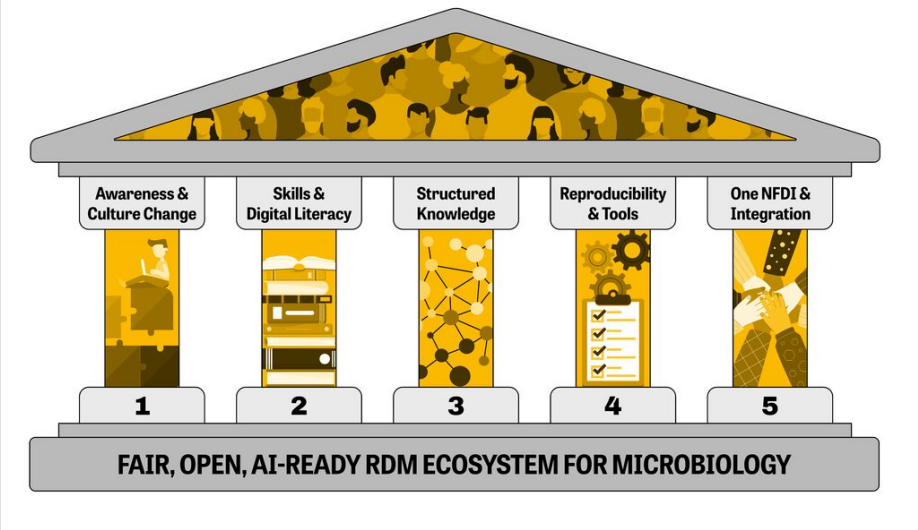

Figure 2.

Depiction of the five key objectives of NFDI4Microbiota in the second funding period.

- Key objective 1: Expand awareness of the importance of FAIR principles, open science, and reproducible research within the microbiology community and drive the cultural change required for their widespread adoption.
- Key objective 2: Equip the community with the skills and digital literacy (data, compute, AI, etc.) required for efficient, data-driven microbial research through a comprehensive training programme.
- Key objective 3: Improve the research process by mobilising, structuring, and linking data, information, knowledge, software, workflows, and models related to microorganisms, using open, community-accepted standards.
- Key objective 4: Make research more reproducible by standardising data processing and analysis and by providing computational tools and databases to translate data into new knowledge.

- Key objective 5: Help to realise the One NFDI vision by building interconnected solutions that transcend disciplinary boundaries.

The success of the NFDI4Microbiota consortium can be effectively measured through a combination of quantitative and qualitative metrics that align with its key objectives. Quantitative measures provide tangible data points that can be easily tracked and analyzed. One crucial aspect is the usage of services, which is tracked by monitoring the number of active users and the frequency of their engagement with the consortium's services. Additionally, the volume of data submitted, processed, and shared through the consortium's platforms serves as a indicator of its impact. Tracking the number of participants in events and workshops along with their completion rates offers insights into the effectiveness of the consortium's educational initiatives. Qualitative measures, on the other hand, offer a deeper understanding of the consortium's impact through subjective feedback and cultural changes. Conducting regular surveys and questionnaires to gather feedback from the community on the usefulness and impact of the consortium's services can provide valuable insights. Collecting and analyzing user testimonials and case reports from researchers who have benefited from NFDI4Microbiota's resources can further illustrate its success.

3 Consortium

- Bielefeld University (UNIBI) is also co-applicant in Basic services for NFDI (Base4NFDI), National Research Data Infrastructure for Biodiversity Research (NFDI4Biodiversity), and Particles, Universe, NuClei, and Hadrons for the National Research Data Infrastructure (PUNCH4NFDI), and participant in Data from PHoton and Neutron Experiments for NFDI (DAPHNE4NFDI), FAIR Data Infrastructure for Agrosystems (FAIRagro), National Research Data Infrastructure for Personal Health Data (NFDI4Health), and National Research Data Infrastructure for the Historically Oriented Humanities (NFDI4Memory)
- European Molecular Biology Laboratory (EMBL) is also co-applicant in German Human Genome-Phenome Archive (GHGA), DAPHNE4NFDI, and National Research Data Infrastructure for Bioimaging Data (Microscopy and Image Analysis) (NFDI4BIOIMAGE)
- Friedrich Schiller University Jena (FSU Jena) is also applicant in National Research Data Infrastructure for Chemistry (NFDI4Chem), and co-applicant in NFDI4Biodiversity and NFDI4BIOIMAGE
- Helmholtz Centre for Environmental Research (UFZ) is also co-applicant in NFDI4Chem, NFDI4Biodiversity, National Research Data Infrastructure for the Material Remains of Human History (NFDI4Objects), National Research Data Infrastructure for Earth System Sciences (NFDI4Earth), NFDI4BIOIMAGE, and FAIRagro
- Helmholtz-Zentrum für Infektionsforschung (HZI) is also co-applicant in National Research Data Infrastructure for Immunology (NFDI4Immuno), and participant in GHGA

- Justus Liebig University Giessen (JLU) is also co-applicant in NFDI4Biodiversity
- German Collection of Microorganisms and Cell Cultures (DSMZ) is also co-applicant in NFDI4Biodiversity, and participant in NFDI4Earth
- University of Marburg (UMR) is also co-applicant in NFDI4Biodiversity and National Research Data Infrastructure for Material and Immaterial Cultural Heritage (NFDI4Culture), and participant in Mathematics Research Data Initiative (MaRDI), NFDI4Memory, and Language- and text-based Research Data Infrastructure (Text+)
- Rheinisch-Westfälische Technische Hochschule Aachen (RWTH) is also applicant in National Research Data Infrastructure for Engineering Sciences (NFDI4Ing), and co-applicant in DAPHNE4NFDI, National Research Data Infrastructure for Catalysis-Related Sciences (NFDI4Cat), NFDI4Chem, National Research Data Infrastructure for Data Science and Artificial Intelligence (NFDI4DataScience), NFDI4Earth, National Research Data Infrastructure for Materials Science and Engineering (NFDI-MatWerk), and PUNCH4NFDI
- ZB MED - Information Centre for Life Sciences (ZB MED, Zentralbibliothek für Medizin) is also co-applicant in NFDI4Health, NFDI4DataScience, and FAIRagro

3.1 Composition of the consortium and its embedding in the community of interest

NFDI4Microbiota benefits from the wide-ranging expertise of its members and will continue its work with the ten (co-)applicants from the first funding phase.

Bielefeld University (UNIBI): With more than 30 years of experience in bioinformatics and life sciences, UNIBI enjoys an excellent international reputation. The Bielefeld Institute for Bioinformatics Infrastructure (BIBI), an interdisciplinary academic department of UNIBI's Faculty of Technology, focuses on the research and service aspects of big-data handling in the life sciences. In collaboration with the German Network for Bioinformatics Infrastructure (de.NBI), it provides comprehensive, state-of-the-art bioinformatics services and training to users from academia and industry involved in basic and applied life science research. Bielefeld Institute for Bioinformatics Infrastructure (BIBI) also promotes cooperation between the German bioinformatics community and international bioinformatics networks such as European life-sciences infrastructure for biological information (ELIXIR). It is involved in several other Nationale Forschungsdateninfrastruktur (NFDI) consortia, including NFDI4Biodiversity, FAIRagro, Data in PLANT research (DataPLANT), and Base4NFDI. **J. Stoye** (Genome Informatics) heads the institute and is the spokesperson for the "Digital Infrastructure for the Life Sciences" graduate school, which forms part of BIBI. **A. Sczyrba** (Computational Metagenomics) heads the Bioinformatics cloud within BIBI, serves as the spokesperson for de.NBI, and is Head of Node of ELIXIR-DE. He also co-organises the community-driven initiative for the Critical Assessment of Metagenome Interpretation (CAMI), which aims to establish and promote standards in computational metagenomics by creating and organising benchmarking challenges.

European Molecular Biology Laboratory (EMBL): EMBL is Europe's flagship institute for molecular biology, life sciences, and biodata. It acts as a hub for microbiome infrastructure through its leading roles in major European and national consortia (including de.NBI, GHGA, NFDI4BIOIMAGE, and the International Human Genome Consortium (IHGC)) and via its subsidiary EMBL-European Bioinformatics Institute (EBI), which hosts essential data resources such as European Nucleotide Archive (ENA). EMBL further advances the field through extensive training and outreach: EMBL's International Centre for Advanced Training (EICAT), its Courses and Conferences Programme, Science and Society Programmes, the European Learning Laboratory for the Life Sciences (ELLS), and the BioIT initiative together offer a wide range of training events and resources for the life science community. **P. Bork**, currently interim Director General of EMBL and Senior Group Leader at its headquarters in Heidelberg, is a world-leading expert on the microbiome. He also serves as Scientific Director of the EMBL-initiated Traversing European Coastlines (TREC) expedition, an integrated, multimodal survey of coastal microbial ecosystems across Europe designed to advance research into planetary health and climate change. Bork's group investigates the microbiome's connections to human phenotypes, diseases, and global ecosystems, developing widely used bioinformatics tools, web services, and databases that collectively receive around one million hits per day.

Helmholtz Centre for Environmental Research (UFZ), Leipzig: The UFZ is one of the world's leading research centres in the field of environmental research. It demonstrates how making sustainable use of our natural resource base—biodiversity, functioning ecosystems, clean water, and intact soils—can benefit both humankind and the environment. Widely regarded as a reliable partner, the UFZ helps policymakers, industry, and the general public to better understand the consequences of human actions on the environment and to develop options for social decision-making processes. The UFZ has strong ties with microbiota research, and several of its departments and groups are participating in NFDI4Microbiota. UFZ's Microbial Data Science group (led by **U. Nunes da Rocha**) works on state-of-the-art research data management in the context of research into microbiomes in environmental samples. U. Nunes da Rocha also coordinates the CoLaborative Multi-domain Exploration of TERRestriAl metagenomes (CLUE-TERRA) consortium that encompasses scientists from 13 institutions and 9 different countries. UFZ's contributions will also include: (i) generating standards for microbiome analysis in pure cultures (Computational Biology and Chemistry, **J. Hackermüller**) and complex microbial communities (Microbial Systems Data Science group, U. Nunes da Rocha); (ii) integrating multi-omics and meta-organisms (Microbiome Biology group, **N. Jehmlich**); Department of MolecularSystems Biology, **M. von Bergen**); (iii) integrating imaging of microbiota at nano scales (led by **H. Stryhanyuk**); and, (iv) bridging human microbiome research to food pathogens, environmentalviromes, and microeukaryotes (led by **J. Schor**).

Friedrich Schiller University Jena (FSU Jena): FSU Jena has a long history as a centre of innovative life science research, particularly in ecology, microbiology, and microscopy. It currently hosts the European Virus Bioinformatics Centre (EVBC), led by **M. Marz**, a

unique international network of virus researchers and bioinformaticians fostering standards and tool development through coordination, training, and organisation of the annual international ViBiom meeting. FSU Jena spearheads the development of VirJenDB, the first umbrella database designed to integrate and harmonise existing virus databases using input from the virus expert community, closing a critical gap in the current bioinformatics landscape. A major research hub is the Excellence Cluster “Balance of the Microverse”, coordinated by **K. Küsel**, which brings together diverse disciplines to study microbial and viral communities in complex ecosystems. FSU Jena serves as a key connector to other NFDI consortia, including NFDI4BIOIMAGE (in collaboration with **T. Figge**), NFDI4Biodiversity (with **B. König-Ries**), and NFDI4Chem (with spokesperson **C. Steinbeck**). These strong cross-consortial ties enable the development of interoperable data standards, shared services, and innovative infrastructure solutions that benefit microbial and viral research alike.

Helmholtz-Zentrum für Infektionsforschung (HZI): HZI is Germany‘s largest public research institution focused on the global challenges of infectious diseases. As a leading centre for innovation-driven infection research, HZI translates fundamental scientific discoveries rapidly and systematically into clinical applications. Its participation in the Centre for Individualised Infection Medicine puts HZI at the forefront of personalised approaches that bring host-pathogen omics data into clinical use. The institution plays a key role in major national research initiatives, including the German Center for Infection Research (DZIF, Deutsches Zentrum für Infektionsforschung) and the National Cohort Study (NAKO). **A. C. McHardy** heads the Department of Computational Biology of Infection Research at HZI. She co-coordinates the DZIF Translational Infrastructure “Bioresources, Biodata and Digital Health” and co-organizes the community-driven initiative for the CAMI, which promotes standards in computational metagenomics through benchmarking challenges. Prof. McHardy serves on the Scientific Advisory Boards of the National Microbiome Data Collaborative (NMDC) in the USA and the Centre for Pathogen Bioinformatics (CPB) in Switzerland, and is a Board Member of the Genomic Standards Consortium (GSC). In addition, she co-chairs the “Microbiome” Community of Special Interest (COSI) within the International Society for Computational Biology (ISCB) and the “Microbiome” track of the Intelligent Systemsfor Molecular Biology (ISMB)/European Conference on Computational Biology (ECCB) conference series.

Justus Liebig University Giessen (JLU): JLU is a leading centre of research on microbiology and infection biology in Germany, focusing strongly on the integration of bioinformatics, veterinary medicine, and translational approaches. The university fosters a dynamic research environment that ranges from basic microbiology to clinical applications and is internationally recognised for its contributions to host–microbe interaction studies. **A. Goesmann** is Professor of Bioinformatics and Systems Biology at the Department of Biology and Chemistry and Vice President for Studies and Teaching. His group develops scalable bioinformatics platforms for secure management and analysis of large-scale microbial genomics, transcriptomics and metagenomics data. Among other research projects, he is a principal investigator in the recently funded German Research Foundation (DFG, Deutsche Forschungsgemeinschaft) Cluster of

Excellence Microbes-for-Climate (M4C) and the LOEWE Priority Program Tree-M. **S. Janssen** leads the research group “Algorithmic Bioinformatics”, which was established in 2019. Their core interests lie in statistical methods for microbiome analyses and combinatorial techniques for sequential data. The group is working towards the implementation of FAIR workflows for microbiome analyses. JLU is also an active partner in NFDI4Biodiversity, contributing to the development of the Research Data Commons (RDC) for cross-consortial data integration and infrastructure alignment. Moreover, the JLU serves as one of the seven national sites of the de.NBI Cloud, providing cloud infrastructure and compute services tailored to life science data needs.

Leibniz Institute German Collection of Microorganisms and Cell Cultures (DSMZ): The Leibniz Institute DSMZ is one of the largest and most biologically diverse microbiological resource centres in the world. In 2023, the new biodata infrastructure DSMZ Digital Diversity was established, integrating scientific databases of fundamental relevance for the life sciences, including the four German Global Core Biodata Resources Bac*Dive*, BRENDA, LPSN and SILVA. A dedicated team of 22 scientists and software engineers are building a common research data infrastructure to significantly improve access to a plethora of curated, standardised biological data. **L. Reimer** heads the database department and leads the DSMZ Digital Diversity consortium. He also heads the development of the Bac*Dive* and StrainInfo databases. Moreover, he is a member of the Central Coordination Unit of de.NBI and of the working group “Digitale Infrastrukturen, Services und Datentracking” in the Alliance of Science Organisations. His research focuses on mobilising and standardising microbial research data to improve accessibility and comparability. **J. Koblitz** leads the Data Integration Group, integrating heterogeneous datasets from different databases via semantic knowledge graphs. Her group also contributes to the database portfolio of DSMZ through the development of Media*Dive* and by augmenting data with AI predictions. **A. H. Scholz** headsthe Science Policy and Internationalisation department. She works at the science-policy interface to ensure that the scientific perspective and empirical data inform new policy measures developed by international UN instruments, including the Convention on Biological Diversity, the High Seas Treaty, and the World Health Organization.

University of Marburg (UMR): The UMR has a long history and an excellent international reputation in microbiology, which has strongly influenced its structural developments over the past decade. **A. Becker** is Professor for Microbial Comparative Genomics and Director of SYNMIKRO, which was jointly established by UMR and the Max Planck Institute for Terrestrial Microbiology (MPIterMIc) in 2010 with a grant from the LOEWE research excellence initiative. SYNMIKRO operates an integrated technology centre comprising topical and infrastructural units in Bioinformatics, Structural Microbiology, Mass Spectrometry, Screening and Automation, Fluorescence Microscopy, and Cell Sorting. This centre is permanently staffed by experts in state-of-the-art technologies and headed by **M. Lechner** of SYNMIKRO’s Bioinformatics unit, who coordinates the Marburg Computing Cluster (MaRC3a) and the Marburg Storage Cluster (MaSC). UMR also participates in BioRoboost, an international collaborative network on synthetic biology standards with a strong focus on microbial systems, and is a co-

applicant in NFDI4Biodiversity (headed by Frank Oliver Glöckner) and NFDI4Culture (headed by Torsten Schrade) which offers a number of synergies. The DFG-funded Cluster of Excellence M4C and Research Training Group 2937 as well the LOEWE Priority Programs Tree-M and RobuCop are based at SYNMIKRO and benefit greatly from the local collaboration with NFDI4Microbiota.

Rheinisch-Westfälische Technische Hochschule Aachen (RWTH): RWTH is one of the leading technical universities in Germany with a focus on engineering and medical sciences. **T. Clavel**, Professor and Head of the Functional Microbiome Research Group at University Hospital of RWTH Aachen (UKA, Universitätsklinikum Aachen), combines (meta)genomic and cultivation approaches to study microbial diversity in the mammalian intestine. His laboratory conducts research into the phylogeny and taxonomy of gut bacteria and defined communities, omics profiling of the gut microbiome in health and diseases (including technical work and training on improving Standard Operating Procedures (SOPs) for sample processing and data analysis), and targeted mechanistic studies in gnotobiotic mice. The team's strengths include cultivation-based microbiome research, as cultured strains and their genomes enable functional studies and the description of novel diversity. T. Clavel participates in several national and international initiativesbringing together experts in microbiology, biotechnology, metagenomics, and (bio)informatics, including: (i) the Euregional Microbiome Centre, a cross-border network (Belgium, Netherlands, Germany) enhancing the visibility of microbiome research in the Euregion (co-founder); (ii) the UNLOCK project (Netherlands), which provides microbiome facilities to the community, including a FAIR data platform (review panel member); (iii) the DFG-funded Priority Program SPP2474 "Illuminating gene functions in the human gut microbiome" (principal investigator); (iv) the Collaborative Research Centres (CRC) 1371 (Freising) and TRR359 (Freiburg) (principal investigator); (v) CRC 1382 at RWTH (Steering Committee Member). In the latter centre, T. Clavel coordinates the core project Functional Microbiota Analysis; this provides standardised wet lab and analysis workflows to the consortium, making it the perfect fit for NFDI4Microbiota. T. Clavel's expertise on microbiome research at RWTH is complemented by expertise in bioimaging (**F. Kiessling**, **M. Grognot**), medical informatics, data protection (**R. Röhrig**), and research data management (Information Technology (IT) Center RWTH).

ZB MED - Information Centre for Life Sciences (ZB MED, Zentralbibliothek für Medizin): Provides numerous services of national and international importance, including PUBLISSO, the PUBLISSO –Repository for Life Sciences (FRL) for Open Access publications, and LIVIVO, the Search Portal for Life Sciences for literature and data discovery. ZB MED engages in a broad range of research activities, spanning from information science to the development of machine learning tools. Furthermore, ZB MED develops and delivers comprehensive training programmes covering a wide array of subjects, including core Open Science practices, the principles of good scientific conduct, programming skills, data management, and knowledge organisation. Spokesperson **K. Förstner** heads the Data Science and Services unit, which develops bioinformatical software, text-mining workflows and knowledge-graph-based solutions. The research group has developed and continues to maintain a variety of services tailored for the

microbiological community. **B. Lindstädt** leads the Research Data Management Group at ZB MED, bringing extensive expertise and experience in user consulting and training. ZB MED is also co-applicant in several other NFDI consortia and participates in multiple Base4NFDI projects (including lead positions for Knowledge Graph Infrastructure for NFDI (KGI4NFDI), Research Data Management Training for NFDI (RDMTraining4NFDI), and Terminology Service for NFDI (TS4NFDI)).

Numerous participants—institutions, individuals, and networks, including several DFG-funded CRCs—also contribute to the consortium. As in the first funding period, NFDI4Microbiota remains deeply rooted in the microbiology research community and closely aligned with its members (co-applicants and participants). Encompassing researchers, infrastructure providers, and data generators, the consortium draws its co-applicants, participants, and other members from key stakeholders in the German scientific community, including universities, the Max Planck Society, the Helmholtz Association of German Research Centres, the Leibniz Association, the Fraunhofer-Gesellschaft, and federal government agencies. NFDI4Microbiota maintains close ties with nearly seven national scholarly societies: Five societies have reiterated their support for NFDI4Microbiota and will continue to help shape its activities: theAssociation for General and Applied Microbiology (VAAM, Vereinigung für Allgemeine und Angewandte Mikrobiologie e.V.), the German Society for Hygiene and Microbiology (DGHM, Deutsche Gesellschaft für Hygiene und Mikrobiologie), the German Society for Parasitology (DGP, Deutsche Gesellschaft für Parasitologie), the German Society for Virology (GfV, Gesellschaft für Virologie e.V.) and the German Society for Mycology (DGfM, Deutsche Gesellschaft für Mykologie). In the new funding period, support will also come from the German-Speaking Mycological Society (DMykG, Deutschsprachige Mykologische Gesellschaft) and the Society for Eukaryotic Microbiology (EuMik, Gesellschaft für Eukaryotische Mikrobiologie e.V.).

These deep connections to the community enable the consortium to address the needs of microbiologists and create demand-driven services. We will continue to systematically strengthen these ties through workshop feedback and through our ambassador programme, which offers bidirectional information pathways aimed mostly, but not exclusively, at early-career researchers. An entire measure is dedicated to maintaining and expanding the bond to our community (M1.3 - Community Engagement), and our Flex Funds programme (M5.3 - Dynamic Adaptation & Growth) will provide the financial support to address needs that have not yet been met.

3.2 The consortium within the NFDI and the national academic research system

NFDI4Microbiota plays a key strategic and operational role in the development and sustainability of the NFDI as a whole. As a consortium at the intersection of microbiology, medicine, agriculture, biotechnology, and environmental science, it operates at the core of multiple research domains and contributes decisively to the horizontal integration of services, standards, and communities within the NFDI. NFDI4Microbiota contributes

substantively to the NFDI through its active participation in all NFDI sections and their working groups, including speaker roles and leadership in section-spanning Task Forces such as Evaluation & Reporting. The consortium acts as a driver of interdisciplinary cooperation, with several of its co-applicant institutions also participating in other consortia (e.g., DataPLANT, FAIRagro, GHGA, NFDI4BIOIMAGE, NFDI4Health, NFDI4Biodiversity, NFDI4Immuno). These overlaps facilitate a continuous and effective exchange of ideas, practices, and solutions across consortia. Joint activities such as shared Use Cases, collaborative Flexible Funds (Flex Funds) (e.g., NFDI4Microbiota-GHGA), and co-hosted training and events (e.g., FAIR data panels, InfraTalks, hackathons such as DataXplorer) are central to our work. These collaborations will be further intensified in the second funding phase through structured mechanisms such as the Biodata Interest Group, BioMed Interest Group, and additional multilateral Memoranda of Understanding (Förstner et al. 2024, Bernard et al. 2025) with related consortia. These efforts are guided by the principle of community-centred integration and sustainability. NFDI4Microbiota addresses a range of cross-cutting topics, particularly in the areas of metadata standards, persistent identifiers, data quality, identity and access management, and training. The consortium is actively engaged in Base4NFDI projects such as KGI4NFDI (knowledge graph infrastructure, led by ZB MED) and RDMTraining4NFDI (training, also led by ZB MED); Data Management Plan for NFDI (DMP4NFDI) (standardised data management plans, as Use Case), PID Infrastructure for NFDI (PID4NFDI) (persistent identifiers, including the StrainInfo Use Case), Identity & Access Management for the German National Research Data Infrastructure (IAM4NFDI) (identity and access management; our Cloud based Workflow Manager (CloWM) and the Aruna data orchestration engine have been selected as Incubator projects for the NFDI Infra Proxy). At the governance level, NFDI4Microbiota is deeply involved in the strategic coordination of the NFDI. Since 2023, the consortium's speaker has served as a member of the Scientific Senate of the NFDI and helped shape the NFDI's overall strategy. Consortium representatives also participate in other key bodies, including the Kommunikationskreis (Communication Circle), where NFDI4Microbiota has co-hosted events and provided significant input on outreach and internal communication strategies. In addition, the consortium helps set strategic priorities through the Task Force Evaluation & Reporting, influencing the evolution of performance indicators and reporting practices across the NFDI. It also contributes to the NFDI's long-term planning through the Task Force Governance and Sustainability. NFDI4Microbiota promotes knowledge transfer both within and beyond the NFDI. It supports community-specific and cross-disciplinary training formats, best-practice dissemination, and the development of guidelines in collaboration with national and international partners. The consortium cooperates with academic stakeholders, research data infrastructures (e.g., de.NBI, ELIXIR's German Node (ELIXIR-DE)), and service providers to ensure long-term accessibility, reproducibility, and reuse of microbial research data. We also contribute to national infrastructure development through alignment with the goals of the National Research Data Strategy and the German Roadmap for Research Infrastructures, and through participation in dialogue on the European Open Science Cloud (EOSC) and other European initiatives.

3.3 International networking

During the first funding period, the consortium built on existing collaborations and forged new ties with key life-science and data-infrastructure initiatives which will be further continued (see Fig. 3 and M1.5 - Connection to International Partners). Member researchers and institutions joined international bodies such as the GSC, ELIXIR, and the Research Data Alliance (RDA). A connection to the EOSC was established, and we were instrumental in discussions that paved the way for the German National node. We established mutual feedback mechanisms and co-developed infrastructures (M4.1 - Computational Infrastructure Operations and Software Components), services (M3.3 - Databases & Data Services), and training (M1.1 - Training & Education) through presentations at meetings, and through booths and workshops at international conferences. Co-organising international conferences and training events enabled synergies with complementary projects such as the NMDC, the Next Generation Sequencing Competence Network (NGS-CN), the EVBC, and The Carpentries. We believe our coordination and community-engagement activities (M1.3 - Community Engagement) have been central to connecting consortium members and therefore the consortium to global communities of researchers, users, and developers. During the second funding period, NFDI4Microbiota will continue to expand and deepen existing international partnerships (ELIXIR, RDA) while pursuing additional strategic alliances in accordance with task h) as stated by the NFDI Expert Committee (NFDI Expert Committee 2023). A key area of collaboration will include the co-development of international metadata standards and services to increase deposition and (re)use of cross-domain data (M2.2 - Data, Metadata & Workflow Standards). We will contribute to establishing the National EOSC Node within NFDI by consolidating our core portfolio of data (M3.3 - Databases & Data Services), services, training materials (M1.1 - Training & Education), and analytical workflows (M3.2 - Analytical Services & Workflows). Our consortium will continue to drive development of the National Node and will contribute tools, services, and expertise to the framework and process supporting the establishment of the EOSC Node (M4.4 - Base4NFDI and EOSC Service Integration). NFDI4Microbiota's work will facilitate international cooperation through emerging and established research, services, training, and infrastructure collaborations in the fields of microbiology.

3.4 Organisational structure and viability

The organisational structure of NFDI4Microbiota (see Fig. 4), as defined in the Consortium Agreement (CA) (3.5.1 Consortium Agreement), has proven effective and has been largely implemented as planned. It enables both efficient coordination and dynamic adaptation, ensuring the successful delivery of strategic goals. The structure is organised as follows:

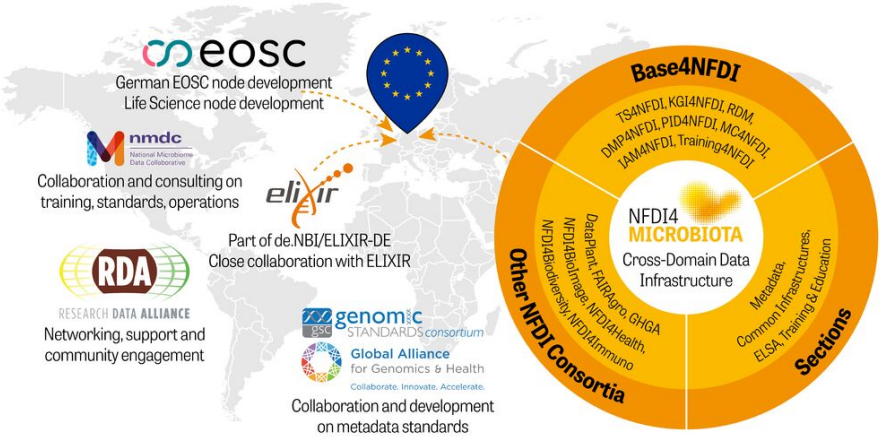

Figure 3.

NFDI4Microbiota embedding in the (inter)national network.

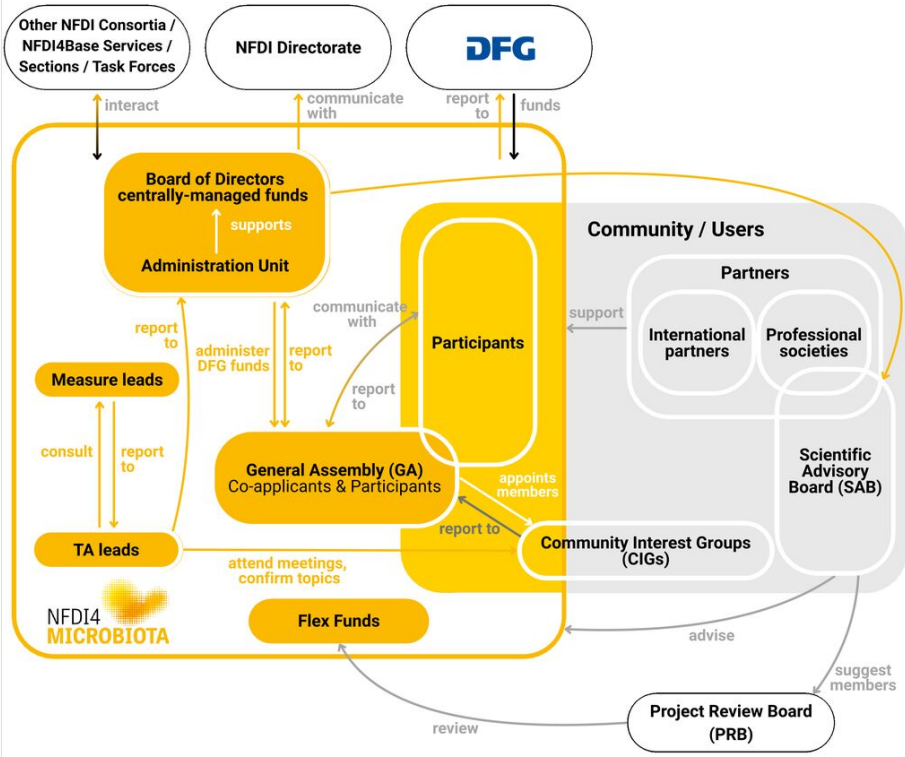

Figure 4.

The consortium structure and interaction with the community as well as the NFDI ecosystem.

Board of Directors (BoD): Although the applicant institution of NFDI4Microbiota is ZB MED, the consortium is coordinated by a modern, dual leadership team consisting of A. C. McHardy (HZI) and K. U. Förstner (ZB MED), both of whom represent NFDI4Microbiota internally and externally. The Board of Directors (BoD), composed of A. C. McHardy (HZI) and K. U. Förstner (ZB MED), provides strategic and operational leadership, coordinates Task area (TA) leads, and ensures proper connections between participants, case studies (formerly Use Cases), and measures. It monitors progress and is responsible for achieving overall project goals. Furthermore, it works with consortium members to resolve issues and supports dynamic adaptation and growth. It ensures compliance with all NFDI requirements, addresses legal, ethical, and contractual matters, and represents the consortium in internal and external settings, including the Consortia Assembly. The BoD meets regularly and maintains direct contact with the NFDI Directorate. It reports to the General Assembly (GA) and Scientific Advisory Board (SAB).

General Assembly (GA): The GA includes representatives from all ten co-applicant institutions (with voting rights) and participants (without voting rights). It is the highest decision-making body for content and organisational matters. The GA monitors TA progress, approves new co-applicants, participants, and project proposals (including Flex and Strategy Fund projects), and ensures alignment with NFDI committees, other NFDI consortia, and national/international infrastructures. It meets at least annually, with pre-distributed agendas, and reports to the BoD.

Administration Unit (AU): The Administration Unit (AU), staffed by project managers, supports the BoD in coordinating the consortium. Responsibilities include internal communication, milestone and deliverable tracking, reporting, quality assurance, and DFG/NFDI administrative tasks. It meets weekly to ensure an agile response and continuity. Evaluating the administrative load as the number of services and partners grows will be key to avoiding bottlenecks.

Task area (TA) Leaders: Each TA is coordinated by a TA Leader who oversees implementation, monitors progress, and ensures timely delivery of milestones within their TA. TA Leaders maintain close contact with Measure Leaders (e.g. to address operational issues) and are responsible for quarterly and annual reporting (on progress and potential risks) to the BoD and GA. Project documentation is maintained using the NFDI OpenProject platform.

Scientific Advisory Board (SAB): The SAB comprises leading scientific experts from academia and industry with experience in microbiology, bioinformatics, and research data infrastructure. It advises on research data management strategies, user needs, emerging global trends, and proposal evaluation, and highlights critical issues. The SAB meets at least once annually and suggests strategic planning ideas to the BoD. Currently, it includes representatives from Würzburg University, Germany (Markus Engstler), the EBI, UK (Robert Finn), Harvard T. H. Chan School of Public Health, USA (Curtis Huttenhower), Joint Genome Institute, USA (Nikos Kyrpides), NMDC, USA (Stanton Martin), Medical University Graz, Austria (Christine Moissel-Eichinger), Laboratory of

Bioinformatics Analyses for Genomics and Metabolism, CNRS, Novozymes, Denmark (Sünje Johanna Pamp), and Hannover Medical School, Germany (Thomas Schulz).

Project Review Board (PRB): Based on SAB nominations, the Project Review Board (PRB) conducts annual reviews of Flex Fund applications to support emerging needs, innovation, and community-driven initiatives aligned with NFDI4Microbiota's mission.

Community Interest Groups (CIGs): Although initially envisioned, User Special Interest Groups (USIGs) were not established during the first funding period. Instead, user feedback and thematic alignment were achieved through alternative formats, including workshops, surveys, training sessions, community events, and participation in scientific networks. This approach combined broad engagement with agile responses to evolving needs. Community Interest Groups (CIGs) will be established during the second funding phase. These will comprise users and developers, who will meet annually.

The consortium is committed to the sustainable delivery of its services through the long-term engagement of its partners and integration with robust infrastructures (e.g., de.NBI Cloud). All software is developed as Free Libre Open Source Software (FLOSS), thus promoting digital sovereignty, transparency and accessibility and further supporting long-term viability and adaptability. Active pursuit of synergies with other NFDI consortia, Base4NFDI projects, de.NBI, and other networks enhances interoperability and resource sharing. NFDI4Microbiota will continue to refine its organisational structure, with a particular focus on consolidating services into sustainable interoperable offerings, fostering cross-consortial collaboration, and strengthening user engagement through structured feedback mechanisms and responsive service development. The chosen structure has enabled NFDI4Microbiota to operate smoothly, adapt to new challenges, and remain aligned with the evolving landscape of research data infrastructure in Germany and beyond. When the overall NFDI structure transitions to the more centralized system proposed by the German Council of Science and Humanities (WR, Wissenschaftsrat) (German Science And Humanities Council 2025) or the RfII (Rat für Informationsinfrastrukturen 2025), we will actively participate in the design, support the transition, and assist in implementing the new governance model.

3.5 Operating model

3.5.1 Consortium Agreement

NFDI4Microbiota operates based on the CA established prior to the start of the first funding period. It relies on a distributed organisational structure consisting of co-applicants and participants, and a centrally established legal framework provided by NFDI e.V. In addition, the consortium maintains close collaboration with the NFDI Directorate and regular communication with other consortia to ensure that the CA remains aligned with evolving governance recommendations and overarching legal and organisational requirements.

3.5.2 Mode of operation and financial model

In its current phase, NFDI4Microbiota provides all its services and offerings free of charge. Any future introduction of user fees will be discussed and decided in coordination with the broader NFDI framework and the other consortia and as part of a transition to a new centralized governance model of NFDI (3.4 Organisational structure and viability). The consortium benefits from substantial in-kind contributions from its members, including personnel, infrastructure, and domain expertise. These contributions were planned from the outset and have proven essential for the delivery of high-quality community services. Key resources such as data storage and computational capacity are distributed across partner institutions and form the backbone of NFDI4Microbiota's operational capabilities. The financial model distributes responsibilities across co-applicants, with dedicated funding allocated for service provision. This distributed approach has demonstrated its effectiveness in ensuring reliable service operation and flexibility in adapting to evolving community needs. To ensure long-term sustainability, NFDI4Microbiota aims to foster institutional embedding of services, strengthen collaboration with NFDI-wide structures, and align with international initiatives such as EOSC-Life. A defined onboarding process and a reserved budget support the inclusion of additional funded and non-funded partners in the future.

3.5.3 Data storage and computing power

NFDI4Microbiota data storage and computing power are provided by three key partners within the de.NBI network: UNIBI, JLU, and EMBL. As active members of de.NBI—now a participant in NFDI4Microbiota—these institutions contribute to the core cloud infrastructure through the de.NBI Cloud, an academic federated cloud computing infrastructure focused on life sciences. The de.NBI Cloud leverages OpenStack technology to offer virtual machines, storage, and pre-configured tools and execution environments that enable researchers to efficiently and safely process large amounts of datasets. It serves as a foundational infrastructure not only for NFDI4Microbiota, but also for several other NFDI consortia, including NFDI4Biodiversity, DataPLANT, FAIRagro, NFDI4BIOIMAGE, and GHGA, and is involved in major European and international projects such as EOSC, HealthyCloud, EGI, and ELIXIR-Converge. As a key infrastructure supporting NFDI4Microbiota, the de.NBI Cloud provides a Cloud currently hosts projects of all sizes and purposes, from developing new services and performing general data analysis to operating core infrastructure for public services. In total, de.NBI has contributed some 630 virtual machines, providing over 6000 cores, 30 TB of Random Access Memory (RAM), 350 TB of storage, an additional 500 TB of object storage, and 5 GPUs. This infrastructure enables several key services (4.4 Services provided by the consortium). For example, CloWM developed by UNIBI provides a user-friendly, standardised, and scalable platform for executing complex data analysis workflows in a reproducible manner (M4.3 - Workflow Execution Platform). Similarly, JLU contributes the Aruna data orchestration engine, which supports secure and distributed data management (M4.2 - Data Storage Platform). As members of de.NBI, UNIBI and JLU ensure that NFDI4Microbiota has access to robust and scalable bioinformatics resources.

Beyond the cloud infrastructure, additional institutional contributions bolster the consortium's technical capacity. The DSMZ hosts the StrainInfo database (M3.3 - Databases & Data Services), HZI maintains the CAMI benchmarking portal (M2.3 - Best Practices for Data Processing), and ZB MED provides hosting for the main web portal (M3.1 - Central Web Portal). These contributions further strengthen the technical foundation and ensure reliable access to critical services.

4 Research Data Management Strategy

4.1 Scientific relevance and quality of the measures

In terms of implementing the planned measures, we are progressing well and largely according to the revised project plan, which was adjusted following the initial budget cuts. Reflecting on the progress of NFDI4Microbiota, we can confidently state that the consortium has achieved the majority of its objectives set for the first funding phase planned for the current moment in time. Moreover, our involvement has extended beyond the initial scope of the proposal. This includes active participation in NFDI sections and task forces, as well as significant engagement with Base4NFDI services, which were conceptualized and launched during the runtime of the current funding phase. Several consortium members have made substantial contributions to the development of Base4NFDI, both in supporting roles and as leaders in specific services, such as RDMTraining4NFDI and KGI4NFDI. Beyond structural and infrastructural achievements, we are beginning to see the impact of our work in the wider microbiology community: After several years of establishing our services, we now observe growing usage and uptake by our target audiences. Qualitatively and thanks in part to the consortium's efforts (Fig. 1), awareness and adoption of the FAIR data principles and open science practices have since improved significantly which is an encouraging indicator of cultural change that aligns with our long-term goals. Clearly, the quantitative degree of adoption and integration into daily research practices is difficult to measure directly. Therefore, the increasing usage of our services and offerings serves as a proxy indicator for this impact. Building on the experiences and insights gained during the first funding phase, we have identified several areas for improvement and strategic refinement. One key lesson has been the importance of balancing service development with sustained community engagement, particularly given the expectedly slow adoption of RDM services by the community, as researchers are often constrained by existing workloads and incentive structures. While our community engagement efforts were already profound during the first phase, we will place even greater emphasis on them going forward (M1.3 - Community Engagement). In doing so, we will diversify our outreach and training activities (M1.1 - Training & Education) in close collaboration with other NFDI consortia. Furthermore, we intend to deepen our partnerships with scholarly societies and research networks, leveraging these connections to enhance visibility, trust, and adoption within the microbiology community. Aside from these actions, we plan to introduce a certification scheme. Labs, institutions, and sequencing centres that adopt NFDI4Microbiota's best practices will be formally recognized for their commitment to FAIR data and metadata,

further incentivising uptake. We also intend to expand our training portfolio through diverse formats, including train-the-trainer workshops, for example, targeting Electronic Lab Notebook (ELN) developers. To assess the impact of our efforts, we track Key Performance Indicators (KPIs) such as participation in training and outreach events. In the next phase, we will align our evaluation criteria with established frameworks (e.g. Minamiyama et al. (2025) and the FAIR Data Maturity Model (FAIR Data Maturity Model Working Group 2020)).

4.2 Metadata standards

NFDI4Microbiota emphasises the adoption, implementation, and harmonisation of metadata standards to ensure FAIR data practices across the microbiology research community (Fig. 5). Metadata standards are a key pillar of the consortium's data infrastructure, enabling seamless data integration, improved findability, and robust interoperability both within the consortium and across the broader NFDI ecosystem. NFDI4Microbiota actively engages with established community-driven metadata standards, including the GSC checklists and extensions for genomic sequences, the Proteomics Standards Initiative (PSI) formats for proteomics, and the Metabolomics Standards Initiative (MSI) specifications for metabolomics. Domain-specific ontologies from the Open Biological and Biomedical Ontologies (OBO) Foundry (e.g., UBERON, ENVO, FMA) are used to annotate metadata fields in these checklists. We are committed to promoting these standards through continuous consultation with researchers and data providers, ensuring their relevance and usability across diverse microbiota-related disciplines. To further improve machine-readability and web discoverability, NFDI4Microbiota promotes the use of Schema.org and Bioschemas, which provide lightweight, web-friendly markup for describing datasets, samples, and biological entities. In addition, the consortium is evaluating and adopting Research Object Crate (RO-Crate) to package datasets together with their rich metadata, provenance, and context in a lightweight, standards-compliant, machine-actionable format. This approach supports better reproducibility and reuse, particularly for multi-modal or complex microbiome datasets. During the initial funding period, NFDI4Microbiota advanced metadata standardisation by assembling community-approved standards into a user-friendly NFDI 4Microbiota Minimal Metadata Standards repository. This repository compiles and evaluates a wide array of technical, biological, and environmental metadata standards tailored specifically for human-, terrestrial-, marine-, plant-, and animal-associated microbiomes, as well as for genomic, metagenomic, amplicon, (meta)transcriptomic, viral, metabolomic, and (meta)proteomic datasets. The consortium contributes to the refinement and development of metadata schemas by participating in national and international working groups and collaborating actively with NFDI-wide initiatives and EOSC-related activities. Such engagement ensures that the consortium's standards reflect scientific advances and evolving community needs. In future, NFDI4Microbiota will enhance interoperability and data reuse by further refining and disseminating metadata standards and the associated tools. An automated metadata submission and validation system will help researchers meet these standards, while existing machine-learning tools for automatic classification and quality control of microbiological data will be expanded to

improve dataset reliability and reusability. Training resources—including educational repositories, workshops (Pauvert and Magel 2024), and tutorials on data and metadata standards—will continue to be developed and disseminated. Metadata stewardship will remain a core Helpdesk service, supported by in-kind contributions from consortium partners with expertise in data curation, metadata management, and ontology development. In the long term, metadata standardisation within NFDI4Microbiota will: (1) enhance interoperability across consortia and domains, (2) foster efficient data reuse and reproducible microbiota research, (3) provide a common language for linking microbiota data with health, environmental, and host-related datasets, and, (4) enable robust integration with international data infrastructures.

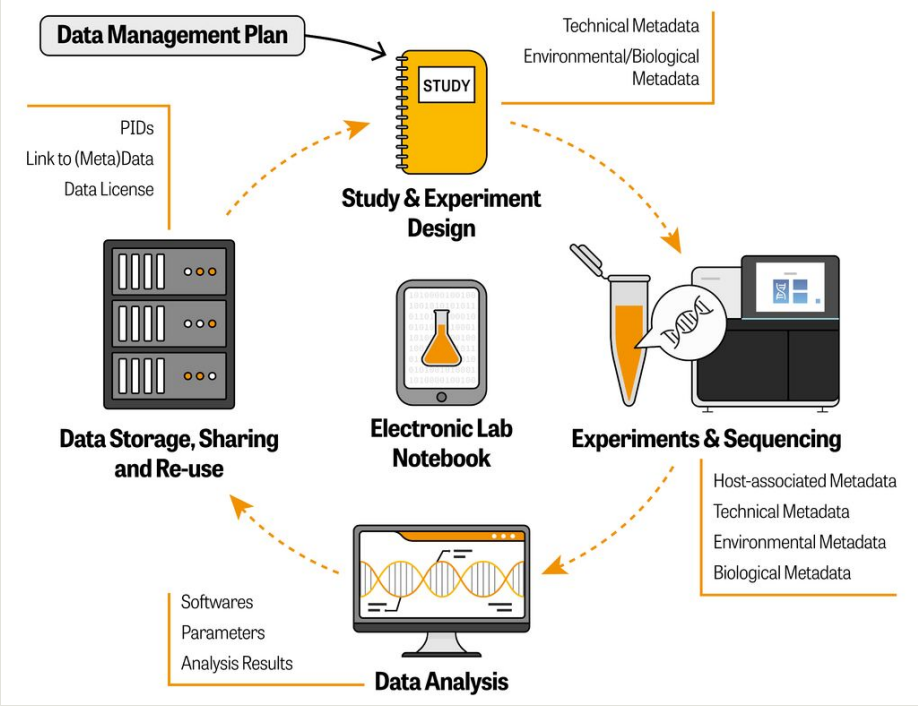

Figure 5.

Metadata collection and standardisation at different stages of the lifecyle of microbiome research.

4.3 Implementation of the FAIR principles and data quality assurance

Our approach centres on capturing metadata either before or during the actual generation of research data, for example during sample sequencing or, ideally, earlier (e.g. using ELNs, M3.5 - Electronic Lab Notebook). We aim to integrate this metadata throughout the entire data life cycle and facilitate submission to community-trusted repositories like the ENA via our services. To ensure seamless data flow from start to finish, we actively engage in all phases of the research process, from the writing of a Data

Management Plan (DMP) to the experimental phase, data generation, analysis, and data deposition and publishing (Fig. 6). We also work on integrating analytical workflows and results to provide comprehensive, cohesive research outputs. A key aspect of our approach is close collaboration with the EBI, part of the EMBL and a partner institution in our consortium. This collaboration enables effective implementation of technical solutions aligned with the latest advances and best practices in bioinformatics. By focusing on these areas, we aim to streamline the research process, enhance data quality and integrity, and ultimately contribute to the advancement of scientific knowledge. In order to assess the FAIRness of our digital assets, including research data and training materials, we intend to use the RDA's FAIR Data Maturity Model (FAIR Data Maturity Model Working Group 2020).

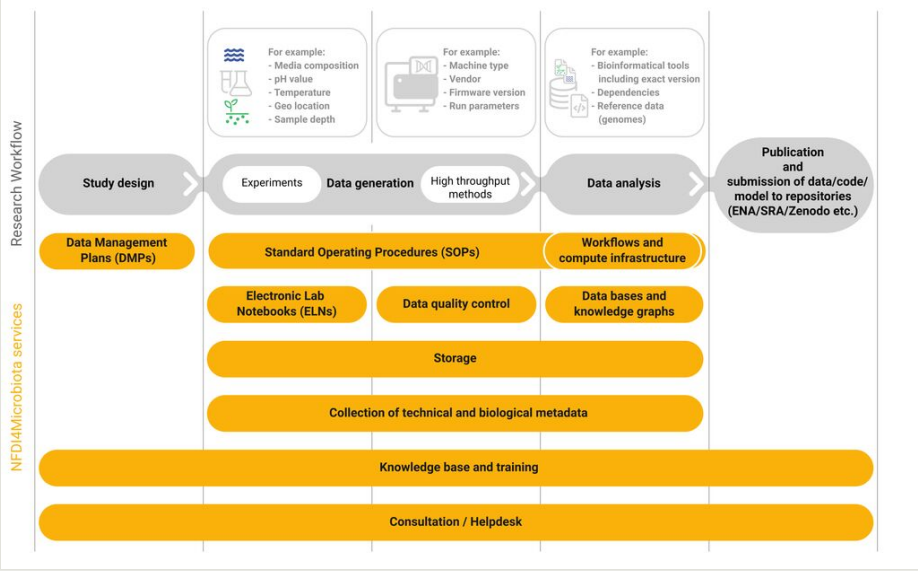

Figure 6.

Overview of the NFDI4Microbiota service portfolio across the research data lifecycle and for an example research project in microbial ecology.

4.4 Services provided by the consortium

NFDI4Microbiota is committed to delivering a comprehensive, quality-assured portfolio of services that address the needs of the microbiology research community, including users and developers, in Germany and beyond. During the first funding period, infrastructure services were developed in conjunction with Base4NFDI to support developers and underpin user-facing services (see M4.1 - Computational Infrastructure Operations and Software Components). This included the release of the CloWM service (M4.3 - Workflow Execution Platform), a cloud-based workflow management system designed to simplify, scale, and improve the reproducibility of omics analyses. As of July 2025, more than 30 production-ready, standardised bioinformatics pipelines, including several nf-core

workflows, can be run from the CloWM web application or embedded into other services via its Application Programming Interface (API). To meet the developer community's needs for scalability, data storage, and syncing, the Aruna data orchestration system (M4.2 - Data Storage Platform) was developed, providing fast, secure, and domain-agnostic data orchestration for user-facing services across several NFDI consortia, and enabling the storage, annotation, and sharing of research data objects according to FAIR principles. Core services such as the Knowledge Base, the HelpDesk (M1.2 - Helpdesk & Support), and the central website (M3.1 - Central Web Portal) provide a gateway to the consortium's user-facing services. Key resources for researchers include the re-developed StrainInfo, now part of the DSMZ Digital Diversity catalogue, to resolve microbial strain identifiers (M3.3 - Databases & Data Services), and the newly launched VirJenDB for curated virus meta(data) and sequences (M3.3 - Databases & Data Services). Through the CAMI benchmarking initiative, NFDI4Microbiota released the widely used CAMI Benchmarking Portal (M2.3 - Best Practices for Data Processing), a platform that enables the reproducible and transparent evaluation of metagenomics tools. The consortium's dynamic project funding via the Flex Funds programme (M5.3 - Dynamic Adaptation & Growth) and showcase activities (M1.6 - Case Studies) enabled the creation, development, and maintenance of innovative tools such as MicroProv for molecular metadata encoding, the establishment of a Qiita instance to support researchers with amplicon project management, and new databases such as MetaTraits, which provides access and annotation services based on a large-scale integration of microbial phenotypic trait information. In line with task d) set out by the NFDI Expert Committee (NFDI Expert Committee 2023), NFDI4Microbiota has focused on extensive training, support and community engagement activities to promote, share, and increase adoption of its services. Its services have featured in training materials, booths, demos, talks, and posters presented at user and developer conferences, workshops, and other training events (M1.1 - Training & Education, M1.2 - Helpdesk & Support), forming a key pillar of our community engagement (M1.3 - Community Engagement) and international networking (M1.5 - Connection to International Partners) activities. The consortium actively monitors service usage by tracking KPIs through the Matomo analytics platform, by capturing metrics such as user numbers, access frequency, and data transfers, and by providing a quantitative feedback mechanism for service developers (M4.5 - Service Monitoring & Reporting). Sustainability and long-term preservation of all services were and remain core priorities as per task g) (NFDI Expert Committee 2023). The consortium actively participates in NFDI's Common Infrastructures Section, contributing to the working group on Long-term access and preservation (LTA) to develop best practices, standards, and sustainable infrastructure. Our strategy involves modular service development and adherence to FAIR guidelines for all research and development outputs, ensuring sustainability and reusability even under uncertain funding conditions (M3.4 - Data Quality & Provenance Services).

The second funding period will focus on enhancing, interconnecting, consolidating, and disseminating but also consolidating existing infrastructure and user-facing services while also developing new services to reach additional communities as per task d) of the NFDI Expert Committee (NFDI Expert Committee 2023). Among the infrastructure

services, CloWM (M4.3 - Workflow Execution Platform) will offer support for multiple workflow languages, adding Snakemake alongside Nextflow, and improve scalability through multi-cloud deployment and real-time workflow monitoring through integration with Base4NFDI services (M4.4 - Base4NFDI and EOSC Service Integration). The Aruna platform will evolve into a fully federated, distributed system to address challenges in scalability, sovereignty, and interoperability (M4.2 - Data Storage Platform), and a flexible ELN ecosystem will be launched (M3.5 - Electronic Lab Notebook). Continued development of the central web portal will provide a focal point for user access to consortium services, including bioinformatics analysis and data management (M3.2 - Analytical Services & Workflows, M2.3 - Best Practices for Data Processing). New user-facing database services such as SPIRE View and Metalog will enable interactive visual exploration of taxonomic profiles, enrich microbiome-related metadata, and link users to the ELIXIR portfolio of services (M1.5 - Connection to International Partners). Within the DSMZ Digital Diversity portfolio, StrainInfo will be maintained and expanded to include the new StrainRegistry, allowing users to register and deposit their own strains directly in a repository. Further, the new MicrobeDive service will extend the DSMZ Bac*Dive* database structure to cover microbial eukaryotes, thereby reaching new user communities directly. Continued development of the virus database VirJenDB will maintain the large, integrated, AI-ready virus datasets and support collaborative metadata standardisation for all viruses (M2.2 - Data, Metadata & Workflow Standards). The CAMI Benchmarking Portal's scope and functionality will be significantly expanded to meet emerging needs across diverse microbiome research communities, establishing best-practice workflows for additional microbial and related domains (M2.3 - Best Practices for Data Processing). Together, the consortium services will develop a unified NFDI4Microbiota knowledge graph (M2.4 - Ontology-driven Knowledge Integration). Consortium, database, and infrastructure services will be seamlessly integrated into training (M1.1 - Training & Education), support M1.2 - Helpdesk & Support, and community engagement activities (M1.3 - Community Engagement), ensuring that existing user communities can continue to shape service development while maximising adoption by new user and developer communities. Collectively, these actions will consolidate services to meet evolving microbiology needs, providing sustainable, user-friendly solutions and support. This will facilitate the efficient storage, utilization, and reuse of microbial search data.

4.5 Impact of changes of external conditions/constraints

As NFDI4Microbiota approaches the second funding period, we must address various external factors that could significantly impact our operations. Although the general needs of the microbiota community remain unchanged, the infrastructure ecosystem and external constraints have evolved substantially, reinforcing the need for NFDI consortia in Germany. As sequencing technologies for microbiome studies continue to advance, the wealth of data generated keeps growing, creating a huge demand for scalable storage systems, a key part of NFDI4Microbiota's remit. Data security has become a critical concern. Stricter personal data regulations, notably the General Data Protection Regulation (GDPR), require projects handling sensitive information to implement robust

measures to safeguard participant privacy. This is particularly relevant for microbiome studies that often involve health-related data and contain personal data. NFDI4Microbiota will continue to work closely with partners such as GHGA and microbiome experts at university hospitals within the consortium to create Secure Processing Environments (SPEs) that meet legal standards yet enable efficient data sharing. We will provide secure storage and clear guidelines for ethical data handling. The integration of AI technologies into microbiome research presents both opportunities and challenges. While AI has the potential to enhance data analysis, automate processes, and drive new insights from complex datasets, it also introduces risks related to data integrity and ethical considerations. To mitigate these risks, NFDI4Microbiota will prioritise the development of robust frameworks for AI governance that ensure compliance with ethical standards while promoting responsible use of AI tools in microbiota research. AI-based methods themselves are heavily dependent on large-scale, high-quality training datasets, including high-quality metadata standards, and these will remain key priorities in NFDI4Microbiota. The potential loss of international databases due to changes in governance and subsequent loss of funding poses a threat. As user needs evolve and datasets grow exponentially, it is essential that our digital infrastructure can accommodate this growth. NFDI4Microbiota helps meet this challenge by enhancing database sustainability, which is essential for long-term success. Collaboration with the EOSC will be crucial in this regard, as it will provide access to shared resources across Europe's research landscape. By aligning our database management strategies with EOSC principles, NFDI4Microbiota can enhance interoperability between various platforms while maintaining local control over sensitive information.

5 Work Programme

An overview of the NFDI4Microbiota work programme is illustrated in Fig. 7.

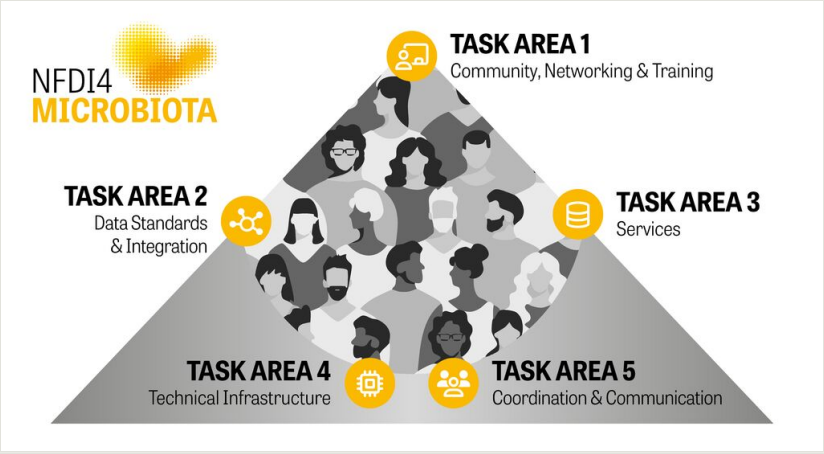

Figure 7.

Overview of the NFDI4Microbiota work programme

Table 9 gives an overview of the task areas (TAs), measures and the responsible co-spokespersons.

Table 9.
Overview of TAs, measures and responsible co-spokespersons

Task Area	Measures	Responsible Co-Spokespersons
1. Community, Networking & Training	M1.1 - Training & Education, M1.2 - Helpdesk & Support, M1.3 - Community Engagement, M1.4 - Connection to other NFDI Consortia, M1.5 - Connection to International Partners, M1.6 - Case Studies	Thomas Clavel (RWTH), Konrad Förstner (ZB MED)
2. From Bench to Repository: Data Standards & Integration	M2.1 - Experimental Practice & Procedure Standards, M2.2 - Data, Metadata & Workflow Standards, M2.3 - Best Practices for Data Processing, M2.4 - Ontology-driven Knowledge Integration, M2.5 - Repository Integration & Data Sharing, M2.6 - Political Consulting & Legal Frameworks	Lorenz Reimer (DSMZ), Alice McHardy (HZI)
3. Services	M3.1 - Central Web Portal, M3.2 - Analytical Services & Workflows, M3.3 - Databases & Data Services, M3.4 - Data Quality & Provenance Services, M3.5 - Electronic Lab Notebooks, M3.6 - AI-based Metadata Imputation, M3.7 - Privacy Sensitive Data Processing & Secure Processing Environments	Peer Bork (EMBL), Manja Marz (FSU Jena)
4. Technical Infrastructure	M4.1 - Computational Infrastructure Operations and Software Components, M4.2 - Data Storage Platform, M4.3 - Workflow Execution Platform, M4.4 - Base4NFDI and EOSC Service Integration, M4.5 - Service Monitoring & Reporting	Alexander Sczyrba (UNIBI), Jens Stoye (UNIBI), Alexander Goesmann (JLU)
5. Coordination & Communication	M5.1 - Project Governance, M5.2 - Project Financial Controlling & Reporting, M5.3 - Dynamic Adaptation & Growth	Konrad Förstner (ZB MED), Peer Bork (EMBL)

Table 10 presents an overview of the tasks set by the NFDI Expert Committee and lists which measures in the working programme addresses them.

Table 10.
How the measures in the working programme contribute to the tasks set by the NFDI Expert Committee (NFDI Expert Committee 2023).

Task set by Expert Committee	Contributing measures
a) (Further) develop research data management for the target group(s) addressed and, leading on from this, identify the core tasks for which long-term funding is required beyond the period of project funding.	M4.2, M5.1

Task set by Expert Committee	Contributing measures
b) Expand inclusion of target community/communities, intensify feedback, increase the use of services beyond the consortium partners themselves, integrate new (sub-)communities where appropriate.	M1.1, M1.2, M1.3, M1.6, M2.1, M2.3, M3.2, M3.3, M3.5, M4.2, M4.3, M4.4
c) Consolidate and expand the relevant structures in order to be able to continuously meet both the evolving requirements of the target groups and the NFDI as a networked infrastructure with a common architecture.	M1.3, M2.3, M3.2, M3.3, M3.7, M4.1, M4.2, M4.3
d) Stabilise, advance and promote a quality-assured service portfolio with regard to the communities addressed to date and also new communities yet to be addressed where relevant.	M1.3, M1.6, M3.2, M3.3, M3.6
e) Expand relevant information and further training programmes in a targeted manner.	M1.1, M1.2, M4.4
f) Implement an organisational model that is sustainable in the long term and ensures a consortium's capacity to engage in participatory processes as well as its sustainable use of human resources.	M1.6, M5.1, M5.3
g) Develop a sustainable model that ensures the continued operation and ongoing funding of services relevant to the target groups as well as NFDI-wide activities in the long term, while at the same time incorporating strategies and measures for innovation.	M1.1, M1.2, M4.2, M4.3, M4.4
h) Expand and deepen local, regional, national and international collaborations so as to increase networking within and outside the NFDI, achieve synergies and increase crossconsortia (re)use of data and services.	M1.1, M1.2, M1.3, M3.2, M3.7, M4.2, M4.3, M4.4

5.1 Task area 1 - Community, Networking & Training

TA-Coordinators: RWTH (lead), ZB MED (co-lead)

TA1 aims to engage a broad community of microbiota research stakeholders and to build a network of mutual support and constructive feedback (Fig. 8). During the first funding period, NFDI4Microbiota successfully provided the community with access to microbiology related data, analysis services, data and metadata standards, and training. Our goal of supporting a broad community and expanding our portfolio accordingly is reflected in the large pool of participating institutions and their researchers who have received training in both basic RDM and specialised microbiome research skills, as well as the continuous onboarding of new participants into the NFDI4Microbiota's pool of stakeholders. A total of more than 190 training events have been held with >2300 participants (M1.1 - Training & Education). Overall, NFDI4Microbiota has been involved in more than 350 conferences and scientific events, ranging from co-organisation of short workshops to comprehensive summer schools delivered through online, on-site, or hybrid models. Building on this solid foundation, future training initiatives will expand into new formats such as on-demand training where research groups can request specific sessions tailored to their interests. The high-quality training and education programme (M1.1 - Training & Education) is designed to meet the needs of researchers at all career stages. In the second funding period, a new framework called "NFDI Comes to You" will

be implemented (M1.3 - Community Engagement); institutions and research consortia across Germany will receive tailored training, ensuring accessibility for all participants as well as a personalised assessment of furtherneeds. This approach will enhance knowledge transfer while identifying specific community needs. Community engagement remains a priority, promoting visibility through targeted outreach activities such as workshops at (inter)national conferences and online tandem talks featuring experts in microbiome research alongside contributions from NFDI4Microbiota service providers. M1.2 - Helpdesk & Support provides critical assistance to scientists throughout the research data lifecycle by involving data stewards in answering queries and by providing curated information on FAIR resources in the NFDI4Microbiota Knowledge Base. Moreover, M1.4 - Connection to other NFDI Consortia will be strengthened through collaborative efforts to co-develop and organise projects and events that address common challenges faced by overlapping target audiences. During the first funding phase, Community Personas — i.e., fictional individuals with a specific role, facing struggles and needs similar to those of our real-life community members— were created to visualise and highlight services that address the needs of our community. Further interconnections between Community Personas, research institutions, and services will result in a more comprehensive portfolio of Services and Interactions. At the international level (M1.5 - Connection to International Partners), strategic links to organisations such as Global Alliance for Genomics & Health (GA4GH), GSC, and EOSC will ensure alignment with global standards while helping to shape policy developments that affect microbiome research infrastructures. Finally, M1.6 - Case Studies will provide targeted opportunities for community members to engage directly with NFDI4Microbiota's service offerings through Community Incubators and Showcases. These actions will not only highlight successful applications but also foster collaboration among users and research institutions to address pressing research questions and identify gaps that require new solutions. Building on the success of the first funding period, we will continue to provide services and improve them based on feedback and experience. During the second funding period, we will enhance support by defining specific roles for data stewards within the consortium and implementing real-time support channels to increase responsiveness. The new features mentioned above and further detailed below in the respective measures will consolidate our interactions with all stakeholders in microbiota research. Task area 1 - Community, Networking & Training is thus a central interface between our consortium and the community, ensuring that we offer the best possible services to maximise their benefits.

5.1.1 M1.1 - Training & Education

Contributors: UNIBI (lead): 54.0 PM, ZB MED (co-lead): 30.0 PM and in-kind, DSMZ: 6.0 PM, EMBL: 20.0 PM, FSU: 13.0 PM and in-kind, HZI: 18.0 PM and in-kind, JLU: in-kind, RWTH: 19.0 PM, UFZ: 15.0 PM and in-kind, UMR: 30.0 PM and in-kind

Goals: The measure will provide high-quality, accessible training for microbiologists and experts at all career stages. We aim to support the community in developing and refining

the skills required to generate, manage, analyse, and share microbiota-related data, paying particular attention to good RDM practices.

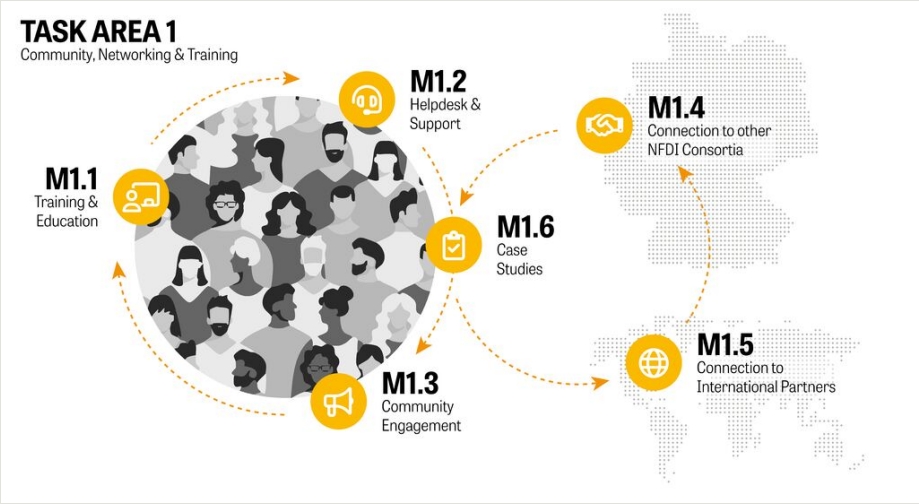

Figure 8.
Overview of Task Area 1

Major accomplishments of first funding period: To meet the diverse needs within the microbiology community, we established a training portfolio covering a broad range of topics. A particular focus was placed on RDM topics (see how we adapted the generic content for microbiology here (Vandendorpe and Lindstädt 2023)), including general principles such as the FAIR Data Principles, as well as discipline-specific strategies that address the unique needs of microbiota researchers (e.g., (Förstner 2024)). Our portfolio also includes technical competencies such as programming skills, domain-specific topics such as data analysis workshops (e.g. Podlesny 2025), and guidance on the effective use of NFDI4Microbiota services (e.g., from M3.3 - Databases & Data Services). In total, we have held more than 190 training events with over 2,300 participants, ranging from short half-day workshops to week-long programmes and summer schools, delivered either online, on-site, or via a hybrid model. We developed a strategy to improve our training by sending post-training evaluations to attendees via a standardised feedback form, allowing us to assess the quality and impact of our training events. To further raise standards, numerous of our trainers completed The Carpentries Instructor Training. A cornerstone of our training and education ecosystem is the NFDI4Microbiota Knowledge Base. The Knowledge Base compiles curated content, practical guidelines, and results from within the consortium. The content is continuously amended and reviewed. During the first funding period, we established regular content and coding sprints to enhance this resource. Organised by our data stewards and open to the community, these sprints have already led to significant enhancements and will continue to be an integral part of our strategy moving forward.

Planned activities for the second funding period: Our traditional workshop-style training events have been well received and will remain a cornerstone of our programme. To broaden accessibility and relevance, we will offer a diverse set of training formats going forward. Alongside traditional live events, we will roll out asynchronous, self-paced formats, including text-based tutorials, videos and self-assessment tools. Modules will be tailored to audiences ranging from beginners to advanced learners. We will also continue to collaborate with consortium services (M3.3 - Databases & Data Services) to co-develop service-specific training materials (e.g., text and video tutorials) and integrate them into our overall programme. Each NFDI4Microbiota partner institution contributes according to their strengths and expertise, allowing us to offer a comprehensive training catalogue (see Table 11). Taking our established Community Personas (see Task area 1 - Community, Networking & Training) and collected feedback into account, we will continuously evaluate and review the training landscape and NFDI4Microbiota's offerings to identify gaps and maintain quality. We have largely adopted The Carpentries training methodology, which provides a solid foundation for effective training. Additionally, we will continue to provide NFDI4Microbiota trainers with Train-the-Trainer opportunities to further enhance their skills. To maximise accessibility and impact, we are committed to publishing training materials as Open Educational Resources (OERs) both inside and outside the consortium (see examples in our Zenodo community). For improved discoverability, we will adopt existing metadata standards for training resources, such as Bioschemas or DALIA. The NFDI4Microbiota Knowledge Base will remain the central hub for information and educational resources. Knowledge Base content requires regular maintenance and moderation, and annual content and coding sprints will be organised by our data stewards. To further boost accessibility and learner engagement, we will produce short, high-quality video tutorials that explain key services, tools, and workflows developed or supported by NFDI4Microbiota. These videos will be published as OERs and integrated into our Knowledge Base and social media channels to reach a broader audience and support self-paced learning. A framework will be made available to help produce our video tutorials and training materials. As part of task e) set by the NFDI Expert Committee (NFDI Expert Committee 2023), we will introduce new structured formats to expand the scope and flexibility of our training. As an extension of the "NFDI Comes to You" initiative (M1.3 - Community Engagement), which will offer tailored on-site or online training, we plan to establish a training-on-demand service where research groups and similar entities can request specific NFDI4Microbiota training sessions from a list of available modules. This approach ensures that training is delivered when and how it is most needed, while also helping us identify high-demand topics. Initially focusing on RDM, the service will expand to cover a wider range of areas based on community interest. We will also implement an Inverted Classroom model. In this format, learners first engage with self-paced online material (text and video) before participating in live sessions designed to address questions and reinforce understanding. Built-in self-assessment tools will help participants track their progress and confidence on the subject matter. A benefit of such an approach is the flexibility learners have to set their own pace. For NFDI4Microbiota, this model offers the additional benefit of gathering frequently asked questions and feedback, which we will use to continually refine and evolve our content. Complete self-learning modules and curated learning paths will be

developed for key topics based on the material produced and feedback acquired in the Inverted Classroom model. Assistance for self-learning modules will be made available through the HelpDesk (M1.2 - Helpdesk & Support). Furthermore, together with M1.4 - Connection to other NFDI Consortia and M1.5 - Connection to International Partners, and in line with tasks b) and h) set by the NFDI Expert Committee (NFDI Expert Committee 2023), we will strengthen links to existing and emerging training resources and services. In addition, we will broaden the consortium's reach to include related communities by collaborating with other life-science-related NFDI4Microbiota consortia (e.g. NFDI4Biodiversity or DataPLANT) and with local and international non-NFDI entities and initiatives such as de.NBI, Landesinitiative Forschungsdatenmanagement Nordrhein-Westfalen (FDM.NRW), Hessian Research Data Infrastructure (HeFDI, Hessische Forschungsdateninfrastruktur), Competence Center Digital Research (zedif, Kompetenzzentrum Digitale Forschung), or Galaxy Training. These partnerships will enhance our training offerings through shared materials, joint events, and cross-promotion. We will also contribute to and utilise training and education services and infrastructure, both within the NFDI framework (e.g., pending their funding status: RDMTraining4NFDI, Data and Language Infrastructure for NFDI (DALIA4NFDI), or DMP4NFDI with their train-the-trainer workshop) and externally (e.g., ELIXIR's Training eSupport System (TESS) platform).

Table 11.

Current and planned portfolio of NFDI4Microbiota training topics (1.–3.). Available Helpdesk experts within the consortium (1.-4.)

Scope	Topics	Experts / Contacts
1. Research Data Management (RDM)	Fundamentals, Open Science & FAIR Principles	FSU Jena, RWTH, UMR, ZB MED
	Electronic Lab Notebooks (ELN)	UMR, ZB MED
	Digital Preservation, Public Repositories - Depositing and Searching	EMBL, ZB MED
2. Software, Infrastructure & Workflows	Programming Fundamentals & Software Development Tools	EMBL, UNIBI, FSU Jena, HZI, RWTH, UMR, ZB MED, JLU
	Machine Learning	HZI, UMR, ZB MED, UNIBI
	Cloud Computing	UNIBI, JLU
	WMSs, Software Frameworks, Containerisation & Orchestration	UNIBI, EMBL, FSU Jena, JLU
3. Domain-Specific Knowledge	Data Generation and Analysis Protocols ((Meta-)genomics, (Meta-)transcriptomics, Multi-omics, ...)	JLU, RWTH, UFZ, FSU Jena, UNIBI, UMR
	Microbial and Viral Bioinformatics, Microbial Pathogenesis & Evolution	HZI, FSU Jena, RWTH

Scope	Topics	Experts / Contacts
	NFDI4Microbiota Services (Aruna, CAMI, CloWM, VirJenDB, …)	JLU, UNIBI, DSMZ, FSU Jena, HZI
4. Community Engagement	Ambassador Programme & Training Topics	EMBL, RWTH, UNIBI, ZB MED
	Case Studies (formerly: Use Cases)	FSU Jena, RWTH, UFZ
	Collaboration & Outreach	FSU Jena, RWTH, ZB MED

Deliverables:

- Continuous offering of training events by all partner institutions
- Yearly evaluation and review of our training offerings
- High-quality video production and social media distribution
- Regular maintenance of NFDI4Microbiota Knowledge Base content
- Annual content and coding sprints for the NFDI4Microbiota Knowledge Base
- Integration of relevant Base services (such as RDMTraining4NFDI and DALIA4NFDI)

Table 12 lists the milestones for M1.1.

Table 12.
Milestones to be achieved in measure 1.1 - Training and education.

Milestone	Month	Description
MS1.1.1	9	First annual meta-evaluation of past training events
MS1.1.2	12	First train-the-trainer event organised
MS1.1.3	12	Launch of the training-on-demand service
MS1.1.4	15	Provide framework for video tutorials and training materials
MS1.1.5	18	Pilot launch of inverted classroom format
MS1.1.6	30	First series of short service tutorial videos released
MS1.1.7	42	Full inverted classroom learning path
MS1.1.8	60	Final version of self-learning modules

5.1.2 M1.2 - Helpdesk & Support

Contributors: ZB MED (lead): 15.0 PM, DSMZ (co-lead): 30.0 PM, UFZ: 10.0 PM, UMR: 30.0 PM and in-kind, FSU: 8.0 PM, JLU: in-kind, RWTH: 6.0 PM

Goals: This measure aims to provide sustained, visible and accessible support to help microbiology researchers make their data FAIR, secure, and reproducible. A core team of

data stewards will continue to offer expert assistance throughout the research data life cycle, e.g., in drafting DMPs (see the NFDI4Microbiota DMP template here (Thölken et al. 2024)), collecting metadata, depositing data and code, and navigating legal issues such as data licensing (see Table 11).

Major accomplishments of first funding period: We established a fully operational Helpdesk, an expanding FAQ resource, and an internal Helpdesk workflow (Vandendorpe et al. 2025). We also co-signed a cross-disciplinary Memorandum of Understanding with other NFDI consortia (Bernard et al. 2025), laying the groundwork for a cross-consortial HelpDesk. Together with M1.1 - Training & Education and M3.1 - Central Web Portal, we launched a comprehensive Knowledge Base covering Frequently Asked Questions (FAQs), best practices, analytical services, and protocols. Protocols are now being migrated to our protocols.io workspace (see M2.1 - Experimental Practice & Procedure Standards for more details).

Planned activities for the second funding period: We will focus on further professionalising and expanding our support for researchers. One goal is to formalise the composition and responsibilities of the Helpdesk team. This includes defining the roles of data stewards within the consortium using frameworks such as Terms4FAIRskills and the Skill4EOSC Minimum Viable Skillsets (Whyte et al. 2024), and appointing a Data Steward Coordinator to oversee coordination and consistency of services. In collaboration with M1.3 - Community Engagement, we will diversify the ways in which we support the community by introducing three new, targeted formats:

- A regular online data steward consultation hour
- An annual online "Ask Anything" session with our core experts
- An annual in-person Data Steward Café to foster informal dialogue

We will also launch an instant-messaging channel for responsive, real-time support. In parallel, we will streamline our internal Helpdesk workflow and, in collaboration with M1.3 - Community Engagement, clearly communicate our services to the community to maximise uptake (e.g., promoting the Helpdesk using posters and business cards at microbiology and RDM conferences). The Helpdesk will also route enquiries about consortium services (M3.3 - Databases & Data Services), offering multiple entry points. If support-request volumes rise significantly, we will introduce a formal ticketing system to ensure efficient processing. We will continue expanding and improving our FAQs, updating it annually to reflect evolving community needs and developments within NFDI4Microbiota. Where appropriate, this FAQs will also be mirrored on broader platforms such as the RDM StackExchange Community to encourage greater visibility. Finally, we will strengthen ties with regional data competence centres, such as Data Literacy Center Rhine-Ruhr (DKZ.2R, Datenkompetenzzentrum Rhein Ruhr) and German Competence Center Cloud Technologies for Data Management and Processing (de.KCD, Deutsche Kompetenzzentrum Cloud-Technologien für Datenmanagement und -verarbeitung) in North Rhine-Westphalia (NRW), to extend our reach and better engage related communities. To enhance our support structure, we will integrate AI-driven first-contact assistance to address common researcher queries, drawing on our extensive

knowledge base and internal documentation. This assistance will be made GDPR-compliant, for instance by using open, self-hosted Large Language Models (LLMs) and fine-tuning them (supported by M4.1 - Computational Infrastructure Operations and Software Components). For more complex issues, the system will automatically route requests to the appropriate experts within the consortium or suggest external specialists within the wider community. This streamlined approach ensures that researchers receive prompt and relevant support while maintaining access to expert guidance when needed.

Deliverables:

- Defined roles and responsibilities for data stewards (e.g., with a survey)
- Formal establishment of Helpdesk core team
- Clear scope of Helpdesk services
- Evaluation report comparing community engagement formats
- Real-time support channel
- AI-based support assistant trained on Knowledge Base

Table 13 lists the milestones for M1.2.

Table 13.
Milestones to be achieved in measure 1.2 - Support.

Milestone	Month	Description
MS1.2.1	6	Election of a Data Steward Coordinator
MS1.2.2	6	Annual plan to promote the Helpdesk at conferences
MS1.2.3	9	Launch and promotion of instant-messaging support channel
MS1.2.4	12	Launch and promotion of new support and outreach formats (i.e. regular consultation hour, annual online "Ask Anything" session, and in-person Data Steward Café)
MS1.2.5	12	Annual FAQ update ahead of the NFDI4Microbiota conference
MS1.2.6	36	AI assistant established

5.1.3 M1.3 - Community Engagement

Contributors: RWTH (lead): 43.0 PM, ZB MED (co-lead): 10.0 PM and in-kind, FSU: 13.0 PM and in-kind

Goals: Community engagement is a key feature inNFDI4Microbiota to raise awareness of useful services and gathers feedback. Its aims are (i) to further boost the visibility of NFDI4Microbiota, its services, and events; (ii) to strengthen interaction with community members; (iii) to keep the consortium agile by continuously integrating user needs; and (iv) to showcase community commitment and involvement through success stories that feature our services.

Major accomplishments of first funding period: We successfully launched an Ambassador Programme that now (mid-2025) includes 65 active ambassadors. The programme relies on early-career researchers and other members of the microbiology community to foster close ties between their institutions and the consortium. Ambassadors receive focused training on RDM and on the use of NFDI4Microbiota's tools, services, and workflows. We also fund their participation in networking events, enabling them to broaden their professional circles and share newly acquired expertise within their home institutions. Ultimately, they serve as knowledgeable local contacts and trusted experts for NFDI4Microbiota users. In addition, we raised awareness and recruited new users by promoting NFDI4Microbiota services through posters, talks, workshops, dedicated sessions, and booths at a total of more than 350 national and international events. Furthermore, we kept the community in touch with important developments through quarterly newsletters, an active news section, a media centre, and three active social media channels (Mastodon, LinkedIn and Bluesky).

Planned activities for the second funding period: We will retain the proven community-engagement framework from the first funding period, ensuring immediate impact and steady promotion of our service portfolio in line with DFG tasks b) and c) (NFDI Expert Committee 2023). To broaden our user base, we will expand the community engagement portfolio in multiple ways to meet the requirements of DFG task b) (NFDI Expert Committee 2023): Based on feedback from the interim report, we have developed Community Personas (i.e., fictional individuals representing target community members who share similar challenges that can be addressed by NFDI4Microbiota's services), thereby underscoring the benefits for important subcommunities. With NFDI4Microbiota core services now more mature, we will establish CIGs to deliver targeted information and foster customer loyalty for each service (see M4.3 - Workflow Execution Platform). Operated largely autonomously, each CIG will maximise impact for its members. To retain current users and attract new subcommunities, the CIGs will be woven into the Ambassador programme and our annual events. A new outreach format, "NFDI comes to you", will be launched in close cooperation with other measures and TAs, especially M1.1 - Training & Education, M1.2 - Helpdesk & Support, M2.1 - Experimental Practice & Procedure Standards, M3.3 - Databases & Data Services, and M3.2 - Analytical Services & Workflows). It will bring tailored, on-site events to institutions across Germany, focusing on major research consortia such as CRCs, Graduate Schools, and Clusters of Excellence. Programmes will be embedded in their specific frameworks (e.g. summer schools and graduate programmes) and adapted to each community's needs, ensuring knowledge is spread throughout the microbiology community and fostering closer cooperation between NFDI4Microbiota and end-users. Pilot events were held during the first funding period (e.g., Biometadata workshops in Freising and Tübingen). During the second funding period we will increase their frequency, topical breadth, and audience reach. This model has already been successfully applied to VirJenDB. Through collaborations with the Annual Meeting of the GfV and the annual Virus Bioinformatics meeting (ViBiom) of the EVBC, one of the largest in Europe, we have already built strong links to the virology and phage research communities, gaining early feedback and visibility for NFDI4Microbiota's M3.3 - Databases & Data Services. Building on this

success, we will be organising dedicated “NFDI comes to you” sessions on virus data infrastructures at key virology hubs such as the Leibniz Institute of Virology (LIV), the UMR, and the University of Freiburg. These events will strengthen community-driven development of VirJenDB, ensuring it aligns with the specific needs and practices of virology researchers. To increase the use of NFDI4Microbiota’s services, in-person and hybrid events will be resumed, and new modes of interaction will be offered:

1. Online tandem talks: A renowned expert from one of our target communities presents state-of-the-art research on a specific microbiome topic, followed by a brief contribution from the consortium highlighting NFDI4Microbiota services and activities that are relevant to that topic.
2. Digital challenges: Short, recurring online sessions that show community members how NFDI4Microbiota can help microbiologists with concrete tasks (e.g., online meeting for “World Clean Up Day” to give scientists tips on data clean-up and storage and to identify suitable challenges for metadata curation). Other measures (e.g., M2.4 - Ontology-driven Knowledge Integration) will be incorporated in the outreach process to maximise the creative potential for of different type of challenges.
3. Champion roles: An extension of the Ambassador Programme for the main users of our services, for members who actively contribute to NFDI4Microbiota, or for individuals who respond to an incentive call (M1.6 - Case Studies). Champions gain additional presentation and networking opportunities, e.g., the chance to present their research to the NFDI4Microbiota community. This in-depth interaction with Champions will foster cross-disciplinary collaboration and ensure that all members of the community can contribute to the development of innovative solutions (M1.6 - Case Studies).
4. Laboratory certification: Labs, institutions, or sequencing centres that implement NFDI4Microbiota’s best-practice recommendations and services will receive certificates to testify their commitment to producing FAIR data and metadata, further incentivising uptake of our services.

NFDI4Microbiota will also continue to interact with the community on general & social media channels. In addition to providing information, this will enable us: to evaluate Community Personas to ensure all relevant communities are included; to tailor content to target audiences and CIGs; to publish educational or entertaining posts that showcase community success stories; and to maintain a feedback loop with community members through surveys and town halls, to name a few examples. To boost national and international visibility, maintain efficient operations, and strengthen community engagement, NFDI4Microbiota will organise annual community events and an international conference in close collaboration with various measures (Task area 5 - Coordination & Communication, M1.5 - Connection to International Partners, M1.4 - Connection to other NFDI Consortia). The international conference, expected to attract 200 participants, will focus solely on community interaction. It will feature keynotes by leading experts and selected short talks and posters from NFDI4Microbiota as well as external contributors, while other community events will also host strategic meetings to

drive new developments. Finally, NFDI4Microbiota shares strong synergies with other NFDI consortia and has signed the “Memorandum of Understanding between NFDI4Biodiversity, NFDI4Microbiota, and DataPLANT” (Förstner et al. 2024), to strengthen ties to the other life and natural sciences domains. These collaborations will result in joint training concepts (M1.1 - Training & Education) and tandem workshop series or co-organised events (M1.4 - Connection to other NFDI Consortia) during the second funding period to address overlapping community needs.

Deliverables:

- Communication channels to collect input from other TAs, relevant measures, and CIGs forstreamlining Public Relation (PR) and engagement strategy (M1.2 - Helpdesk & Support, Task area 3 - Services, Task area 2 - From Bench to Repository: Data Standards & Integration, Task area 4 - Technical Infrastructure)
- Publicly available calendar instance and press-release kits

Table 14 lists the milestones for M1.3.

Table 14.
Milestones to be achieved in measure 1.3 - Community Engagement

Milestone	Month	Description
MS1.3.1	3	Resume engagement cycle with events, series of talks, news, and social media updates
MS1.3.2	9	Create framework for new concepts and initiate digital challenges, "NFDI comes to you", incentive programs of M1.6 - Case Studies, CIGs
MS1.3.3	9	Launch a recurring community survey and first virtual town hall
MS1.3.4	12	Create framework for new engagement strategies: "focus topics", community needs/ personas including evaluation for online and PR strategy
MS1.3.5	18	Expand concept of Ambassador Programme & Championship
MS1.3.6	12, 24, 48	Organise annual community events (including bi-annual cooperation events with relevant CRCs and NFDI consortia)
MS1.3.7	36	Organise International Conference
MS1.3.8	48	Establish lab certification process

5.1.4 M1.4 - Connection to other NFDI Consortia

Contributors: ZB MED (lead): 5.0 PM, EMBL (co-lead): in-kind, FSU: in-kind, JLU: in-kind, RWTH: in-kind

Goals: To strengthen our interactions with other NFDI consortia, sections, task forces, and Base4NFDI services to align activities, avoid duplication, and create cross-cutting synergies. Rather than focusing on technical integration (covered under M4.4 -

Base4NFDI and EOSC Service Integration), this measure focuses on communication, coordination, and mutual knowledge transfer.

Major accomplishments of first funding period: NFDI4Microbiota is now deeply embedded in the NFDI ecosystem (see 3.2 - The consortium within the NFDI and the national academic research system), contributing to all sections, multiple task forces, and several Base4NFDI services. We also participate in numerous interest groups that build synergies around domain-specific topics, standards and practices.

Planned activities for the second funding period: Our participation in relevant NFDI-wide structures will continue, including contributions to NFDI section working groups and task forces as well as to Base4NFDI projects. We will continue to engage with members of the Biodata (Förstner et al. 2024) and BioMed interest groups to ensure that domain-specific developments are coordinated and broadly communicated. As part of this measure, we will also continue to promote our technical solutions to other consortia and actively explore their portfolio to identify complementary services that could be of use to our community. If recommendation to restructure the NFDI into a more centralised system are adopted in the upcoming agreement between the federal government and the federal states for the post-2028 NFDI set-up (German Science And Humanities Council 2025), we will actively support this transition, particularly through close collaboration with other life science consortia.

Deliverables:

- Manage the contribution to the sections and their working groups and task forces
- Workflow for sharing information on training events with other life science consortia with the aim of co-hosting such events
- Interaction with BioData and BioMed interest groups

Table 15 lists the milestones for M1.4.

Table 15.
Milestones to be achieved in measure 1.4 - Connection to other NFDI Consortia

Milestone	Month	Description
MS1.4.1	1	Overview of interaction responsibilities created
MS1.4.2	12	BioData interest group workshop conducted
MS1.4.3	18	BioMed interest group workshop conducted

5.1.5 M1.5 - Connection to International Partners

Contributors: EMBL (lead): 18.0 PM and in-kind, FSU (co-lead): 8.0 PM and in-kind, JLU: in-kind, RWTH: in-kind, UFZ: 7.5 PM and in-kind, ZB MED: in-kind

Goals: This measure will maintain and strengthen strategic partnerships with key international initiatives in life-science data sharing and infrastructure development, and to establish new ones. By leveraging our ties with GA4GH, EMBL-EBI, and especially the EOSC, we will ensure alignment with global standards and actively contribute to shaping a sustainable, interoperable research data ecosystem beyond the NFDI.

Major accomplishments of first funding period: During the first funding period, member researchers and institutions joined national and international bodies such as the GSC, ELIXIR, and the RDA. We established ties to the EOSC and contributed to plans for a German EOSC node. We also set up collaborations and two-way feedback channels on training and infrastructure services with parallel projects, including the NMDC, the NGS-CN, VIROINF, EVBC as well as The Carpentries and exchange with the Galaxy community was initiated.

Planned activities for the second funding period: For the second funding period, key areas of collaboration include co-developing international metadata standards (Schema.org/Bioschemas and GSC, 4.2 - Metadata standards) with the NFDI BioData and BioMed interest groups, and merging a core portfolio of services and training for the FAIRification of resources, analytical workflows, and training materials (CloWM, nf-core community, EOSC). EMBL, as a core project partner, provides direct connections to both EBI and EOSC governance structures. NFDI4Microbiota participants are involved in regular coordination meetings with the European Commission, EOSC Association, and European Strategy Forum on Research Infrastructures (ESFRI) bodies. These interactions provide early access to policy developments and allow direct input into initiatives such as the emerging EOSC Life Science Node. Strategic connections from the first funding phase —such as with the GSC, NMDC, and ELIXIR— will be maintained through cross-representation at conferences and meetings, joint proposals, project collaborations, and co-development of resources (see M3.3 - Databases & Data Services and M4.1 - Computational Infrastructure Operations and Software Components). This measure will coordinate our contributions, identify concrete integration points (e.g. service onboarding, metadata alignment, training), and establish a long-term strategy to sustain these connections beyond the project's lifetime.

Table 16 lists the milestones for M1.5.

Table 16.
Milestones to be achieved in measure 1.5 - Connection to International Partners

Milestone	Month	Description
MS1.5.1	6	Formal liaison established with EOSC Life Science Node team; definition of integration roadmap and alignment points
MS1.5.2	18	Active participation in at least one EOSC or GA4GH working group; documented contribution to international interoperability or metadata standards

5.1.6 M1.6 - Case Studies

Contributors: RWTH (lead): 35.0 PM, FSU (co-lead): 3.0 PM, DSMZ: in-kind, UFZ: 7.9 PM and in-kind, ZB MED: in-kind

Goals: This measure will empower the community to draw on NFDI4Microbiota's services to facilitate their own research and to identify and fill gaps in the service architecture. This interaction will be enhanced by incentivising community contributions to existing services and/or solving their research questions in new ways. The different types of Case Studies described below (Community Incubators, Showcases, Incentive Calls) will offer "all-hands-on-deck" support for scientific projects by adopting existing or newly developed solutions into NFDI4Microbiota's service portfolio. This will allow NFDI4Microbiota to reach new (sub-)communities, as recommended by the NFDI Expert Committee (task b) (NFDI Expert Committee 2023).

Major accomplishments of first funding period: During the first funding period, we established nine diverse Showcases (formerly referred to as "Use Cases"). These delivered tools, guidelines, and workshops on multi-omics preprocessing recommendations, RDM portals for metaproteomics data, taxonomic assignment of novel bacterial taxa, traceable strain information, and curated (meta)databases for microbial (meta)data associated with soil, human, animal, virus, and plant samples. Developed by partners, participants, and external contributors, these Showcases aimed to: (i) test and fine-tune NFDI4Microbiota solutions for use in real scientific projects or as reusable tools; (ii) pool stakeholder expertise to create sophisticated solutions that that are readily adaptable to other research projects; and (iii) reveal new community needs for which the consortium could introduce additional measures.

Planned activities for the second funding period: To provide more targeted solutions that address diverse community needs, a new "Case Studies" framework will be established in the second funding period with three levels of interaction: 1. Community Incubators, 2. Showcases, and 3. Incentive Calls. Varying in objective, duration, and intensity, these Case Studies will be deeply embedded in the work programmes of selected services (see Task area 3 - Services and Task area 4 - Technical Infrastructure) and will draw on existing interest from CIGs.

1. Community Incubators will issue bespoke, time-limited calls inviting community members to test and refine specific NFDI4Microbiota services. Working in collaboration with TA and Measure Leads (especially Task area 3 - Services), we will identify needs on both the community and consortium sides and co-develop solutions. Should in-kind contributions from community and consortium members be insufficient for more labour-intensive Community Incubators, we will propose these as tailored Flex Fund Projects (M5.3 - Dynamic Adaptation & Growth) to increase capacity.
2. Showcases will build on our first-period Use Case strategy by demonstrating how NFDI4Microbiota core services can be used to solve research questions and produce FAIR data. Unlike Community Incubators, Showcases offer a long-term

window of interaction between the community and NFDI4Microbiota. While Community Incubators take a multilateral approach that considers various aspects and drives strategic development of translational features, Showcases adopt a bilateral approach that highlights services through more direct community interaction and engagement.

3. Incentive Calls—developed with M1.3 - Community Engagement and Task area 3 - Services— will encourage targeted contributions to fill specific gaps in the NFDI4Microbiota framework, identified either by our measures or via Showcases and Community Incubators. Examples are calls for data curation or data population, for test users to submit data to ENA (M2.5 - Repository Integration & Data Sharing), for ELN feature development (M3.5 - Electronic Lab Notebook), and for other ideas that will anchor NFDI4Microbiota's core services in the community using relevant tools.

To visualise and update the many interactions and dependencies within NFDI4Microbiota's portfolio that lay the foundations for Community Incubators, Showcases, and Incentive Calls, we will create a Services and Interactions Map that shows service-to-service, service-to-infrastructure, and user-to-service connections. This map will help to promote Community Engagement (M1.3 - Community Engagement) and Flex Fund calls (M5.3 - Dynamic Adaptation & Growth) while revealing critical infrastructure gaps that need attention. The new set of NFDI4Microbiota Showcases and Community Incubators will be woven into the working programme and developmental framework, allowing us to establish and adapt demand-driven services (see for example M3.2 - Analytical Services & Workflows and M3.3 - Databases & Data Services). Future Case Studies will include one based on the TREC expedition, which generated multi-omics data and expanded the pool of multi-omic workflows on CIoWM (M3.2 - Analytical Services & Workflows). Furthermore, the need for technical and M2.6 - Political Consulting & Legal Frameworks for establishing M3.7 - Privacy Sensitive Data Processing & Secure Processing Environments was an important consideration identified in the first funding period. Bringing the key stakeholders to identify new needs of our community members is at the core of the Case Studies.

Deliverables:

- Communication of Case Study developments in collaboration with M1.3 - Community Engagement and Task area 5 - Coordination & Communication
- Visualisation of stakeholder interactions and connection points resulting in a comprehensive Service and Interaction Map in cooperation with M2.4 - Ontology-driven Knowledge Integration and Task area 5 - Coordination & Communication with input from each measure

Table 17 lists the milestones for M1.6.

Table 17.
Milestones to be achieved in measure 1.6 - Case Studies

Milestone	Month	Description
MS1.6.1	6	Framework established for each Case Study format
MS1.6.2	12	Regular dynamic assessment of framework and onboarding process
MS1.6.3	18	Regular implementation of Case Studies tailored to the established framework
MS1.6.4	24	Framework established to highlight key interactions between Case Study components
MS1.6.5	36	Detailed draft of Services and Interactions Map and regular updates

5.1.7 Overall schedule, contribution, risks and dependencies

Fig. 9 illustrates the overall schedule for TA1.

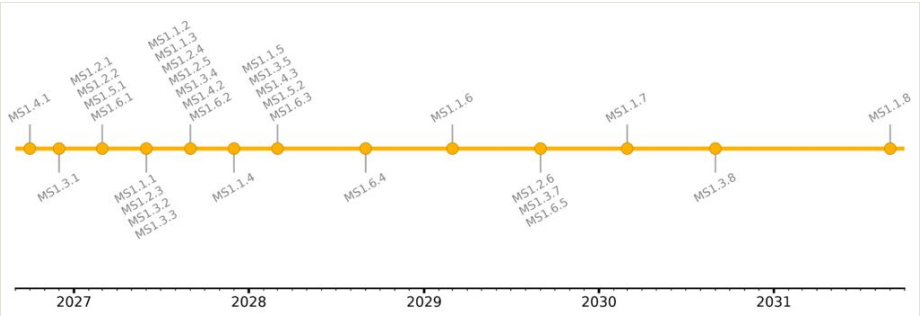

Figure 9.
Overall schedule Task Area 1

Contributions: Community engagement is central to NFDI4Microbiota and is reflected in Task area 1 - Community, Networking & Training. All co-applicants contribute to this TA, with RWTH, ZB MED, UMR, and UNIBI leading outreach and training efforts. RWTH coordinates engagement strategies (M1.3 - Community Engagement), and M1.6 - Case Studies, in particular, helps to create collaborative scientific projects that bring together stakeholders such as institutions, individual researchers, international initiatives, and, crucially, NFDI4Microbiota's service developers and providers, who support community activities as set out in the measures Task area 3 - Services, Task area 4 - Technical Infrastructure and Task area 2 - From Bench to Repository: Data Standards & Integration. The contributions of other co-applicants average around 50 person-months across nine institutions, underscoring their shared commitment to a robust network. Additionally, in-kind contributions include existing training programmes on RDM principles and workshops on microbiological data analysis skills, which will be refined as the measures advance; new content will be developed in cooperation with other TAs, international partners, fellow NFDI consortia and Base4NFDI services to meet emerging needs.

Dependencies: TA1's success is intrinsically linked to Germany's broader microbiota-research landscape. NFDI4Microbiota acts as the interface between the co-applicants' expertise and a heterogeneous community that ranges from wet-lab bacteriologists to bioinformaticians and software engineers. Clear communication between the consortium and these communities not only enhances engagement, but also keeps the feedback loop active for continuous improvement. This TA relies heavily on collaboration with other task areas within NFDI4Microbiota; for example, training developed in M1.1 - Training & Education will directly inform best practices defined in M2.2 - Data, Metadata & Workflow Standards, which are harmonised with international efforts through M1.5 - Connection to International Partners and adopted by services and databases in Task area 3 - Services. These connections keep TA1 resilient to external pressures and internal adjustments within NFDI4Microbiota. To broaden the European reach of our events and training, we will use platforms such as ELIXIR's TESS.

Risks of implementation and mitigation approaches: During the first funding period, NFDI4Microbiota forged close collaborations and set up services for target communities working on metaproteomic analysis, novel strain and functional taxonomy analysis, virology, metagenomic analyses, and environmental sample collection with accompanying metadata. To retain these communities, we will co-organise events with thematically relevant CRCs as part of the "NFDI comes to you" programme. (1) One potential risk is that new target communities may not actively engage with NFDI4Microbiota because they deem our services irrelevant or remain unaware of them. We will mitigate this by continuing to develop services proactively and collaboratively with our community. During the first funding phase, we received positive feedback and suggestions for improvement from surveys conducted at multiple events. To keep engagement with specific user groups high, our targeted outreach plans will employ personalised communication based on Community Personas and Showcases. (2) Another risk is training demand outstripping capacity; some computational-method courses have been double or triple-booked. We will relieve this pressure through inverted classroom formats and additional self-paced materials that foster participants' self-agency. (3) Because M3.3 - Databases & Data Services and M3.2 - Analytical Services & Workflows embed outreach in their core milestones, one risk is that services could become decoupled from the common vision and outreach efforts established for the consortium as a whole. To mitigate this risk, various survey and reporting procedures will be agreed and implemented with the TA leads and Task area 5 - Coordination & Communication to keep communication consistent and development aligned. (4) At the international level, NFDI4Microbiota must continually refine and align its services with emerging regulations and evolving trends that strongly affect the community. To inform ongoing discussions, Task area 1 - Community, Networking & Training will engage closely with work done by M2.6 - Political Consulting & Legal Frameworks. One risk is that target community members may initially be overwhelmed by new requirements within their institutional working practices, but we are confident that NFDI4Microbiota members' extensive experience and networks will enable us to jointly identify suitable long-term solutions for community members.

5.2 Task area 2 - From Bench to Repository: Data Standards & Integration

TA-Coordinators: DSMZ (lead), HZI (co-lead)

TA2 lays the foundation for FAIR, open and AI-ready data, promoting long-term reusability and integration within and beyond the NFDI4Microbiota consortium by strengthening interoperability and standardisation (Fig. 10). In the first funding period, the TA compiled and published current standards for microbiological procedures and metadata, launched a tool for benchmarking bioinformatics software, collected best-practice recommendations for bioinformatics workflows, and uploaded workflows that meet these recommendations to the CloWM platform. In the second funding period, the TA will continue to promote the adoption of RDM practices by implementing community-accepted standards and harmonised metadata, thereby improving the interoperability, reusability, and comparability of microbiological research data. Our goal is to enable seamless data integration across disciplines, institutions, and infrastructures, thereby empowering researchers to build on one another's work and create added value from existing resources. Support for data producers (M2.1 - Experimental Practice & Procedure Standards) will include publication of SOP standards across various subcommunities and the establishment of a video-learning platform for cultivation methods, both connected to the NFDI4Microbiota web portal. Additionally, a new registry for cultivation media will be developed to improve FAIR sharing of cultivation protocols. To further enhance data interoperability and sharing, a measure will be introduced to define, promote, and support community-driven standards for data, metadata, and analysis workflows in microbiological research (M2.2 - Data, Metadata & Workflow Standards). This includes the development of a novel microbiology-specific metadata schema using standardised ontologies, which will ensure full alignment with FAIR principles and technical interoperability across various disciplines and platforms. This will be backed by extensive documentation, training, and outreach. Researchers will receive targeted support in choosing bioinformatics tools that reliably meet their needs (M2.3 - Best Practices for Data Processing) thanks to rigorous benchmarking of methods on complex, realistic datasets and clear decision-making guidelines. To improve the organisation and discoverability of microbiological data, M2.4 - Ontology-driven Knowledge Integration will lay the groundwork for its transformation into AI-interpretable knowledge. This will be driven by the development of controlled vocabularies, ontologies, and schema mappings that align with existing efforts from ELIXIR, the INSDC, and NFDI consortia. Consistent with the Knowledge Graph Alliance's XAIR (eXplainable-AI-Ready) concept, annotating datasets with ontologies is essential for maximising the power of interpretable machine-learning methods. To enable early, standards-compliant preservation and sharing of research data, we seek to streamline data flow from generation to repository (M2.5 - Repository Integration & Data Sharing). Data-generating facilities, such as sequencing and proteomics labs, will be directly integrated and supported in the data collection and submission process. Researchers will receive assistance in submitting micobiology (meta)data to repositories through specific ready-to-use data templates, validation tools, and guidance documents tailored to the needs of the microbiology community. Close collaboration with Task area 3 - Services and Task area 4

- Technical Infrastructure will ensure that the developed standards are directly embedded in workflows, databases, and submission tools. Lastly, this task area will address the increasing importance of legal and political frameworks in microbiological research (M2.6 - Political Consulting & Legal Frameworks) by actively participating in policy debates, raising awareness in the community, and aiding researchers in handling the complex legal obligations that can arise in the handling of research data.

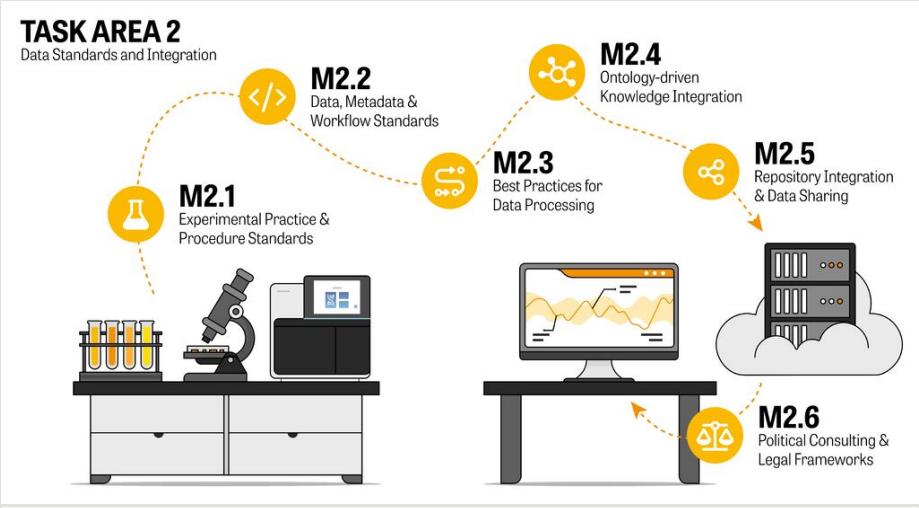

Figure 10.

Overview of Task Area 2

5.2.1 M2.1 - Experimental Practice & Procedure Standards

Contributors: DSMZ (lead): 60.0 PM, RWTH (co-lead): 44.0 PM, EMBL: 10.0 PM, FSU: in-kind

Goals: Variations in wet-lab practice – from sample collection to experimental conditions – can profoundly affect results. This measure will provide comprehensive support for data producers across the microbiota community, from sequencing centres to microbiology labs, by publishing SOPs and developing tools for FAIR data sharing. It will bring us closer to the end-user community of data producers (e.g., new PhD candidates) and help us understand and address their needs. By combining protocol expertise (RWTH, EMBL) with standardised data infrastructures and registration platforms (DSMZ), we are ideally positioned to support experimental workflows from bench to database.

Major accomplishments of first funding period: In the first funding period, we assembled a set of SOPs for experimental research, combining in-house protocols with curated SOPs from high-impact, peer-reviewed journals, and published the collection on protocols.io and the NFDI4Microbiota Knowledge Base.

Planned activities for the second funding period: The set of SOPs compiled during the first funding phase requires constant updating and further expansion. We will focus on omics-informed cultivation methods (Clavel et al. 2025) during the second funding period due to the growing community interest in the use of bacterial isolates. It is essential to strengthen this expertise in isolate-based microbiota research within NFDI4Microbiota to meet the current and future needs of an expanding research community (e.g. SPP2474 starting in September 2025). RWTH and DSMZ have long-standing collaborations in this field. To guide the community towards FAIR experimental practices, we will pursue two major goals: 1.) Maintaining high quality SOPs: We aim to leverage existing platforms such as protocols.io to feature community-endorsed SOPs in a central registry while advocating for users and their SOPs to remain FAIR and open in the face of future changes to the publication process and landscape. Acting as a catalyst, we will support researchers in publishing their protocols in compliance with metadata standards. Leveraging our strong connections to CRCs, Priority Programmess (SPPs), research clusters, and ambassadors, we will gather high-quality SOPs from diverse domains. A key aspect will be the development of an incentive strategy to motivate contributions, especially during targeted calls and knowledge base sprints. To make SOPs FAIRer, we will evaluate how to establish machine-actionable, interoperable, and accessible SOPs through the NFDI4Microbiota website (M3.1 - Central Web Portal), link them to ELNs (M3.5 - Electronic Lab Notebook), and supply templates to help users create new protocols from scratch. In the longer term, we will publish a roadmap that connects SOPs across experimental workflows such as sequencing or proteomics, enabling the community to harmonise procedures and convert existing SOPs into FAIR formats. 2.) Develop standards and tools for FAIR reporting of microbial cultivation media: We will expand and further develop the Media*Dive* platform as a central hub of standardised microbial cultivation media and growth condition data (Koblitz et al. 2023). To boost reproducibility and support early-career scientists, we will introduce a video-based learning platform to teach cultivation techniques. This will include interviews with cultivation experts from DSMZ, RWTH, and other institutes within the NFDI4Microbiota community. The resulting content will be integrated directly into the Media*Dive* interface and made available in the NFDI4Microbiota Knowledge Base. To support biotechnological Use Cases, we will extend the Media*Dive* schema to accommodate fermenter-based cultivation and yield reporting. Furthermore, we will transform the existing Medium Builder tool into a community-driven platform for the design and registration of cultivation media. Researchers will be able to register for free, create new media recipes, manage access permissions (e.g., public or closed), share drafts in collaborative or view-only mode, modify existing media, add strains and growth data, and comment on cultivation success. Upon publication, all media will receive unique persistent identifiers and DOIs, ensuring citability and long-term accessibility. Close collaboration with StrainRegistry (M3.3 - Databases & Data Services) will allow for joint registration of strains and associated media, thereby facilitating the submission of media recipes to biological resource centers when strains are deposited (Koblitz et al. 2024). To ensure seamless workflows, Media*Dive* will be integrated with ELNs (M3.5 - Electronic Lab Notebook) to enable automated reuse and traceability of cultivation conditions. In addition, we will evaluate existing vocabularies and ontologies for cultivation media,

ingredients, and lab equipment – and develop new ones where necessary – building on established resources such as ChEBI and GMO(M2.2 - Data, Metadata & Workflow Standards). All resulting data and metadata will be fed into the NFDI4Microbiota knowledge graph (M2.4 - Ontology-driven Knowledge Integration), enabling semantic interoperability and supporting downstream AI applications.

Deliverables:

- Functional release of Media*Dive* with enhanced data schema for bioindustry
- First release of MediaRegistry as a media curation interface
- Full release of MediaRegistry for cultivation media submission and data sharing
- Functional protocol market with best practices and SOPs from a wide range of microbiology-related experiments

Table 18 lists the milestones for M2.1.

Table 18.
Milestones to be achieved in measure 2.1 - Experimental Practice & Procedure Standards

Milestone	Month	Description
MS2.1.1	2	Concept and technical roadmap for Media*Dive* extensions and SOP marketplace
MS2.1.2	10	Extended Media*Dive* schema
MS2.1.3	24	First expert interviews recorded and integrated into the NFDI4Microbiota Knowledge Base
MS2.1.4	24	First functional prototype of MediaRegistry including account system and draft sharing
MS2.1.5	36	DOI registration for cultivation media and lab protocols
MS2.1.6	48	ELN integration prototype completed and tested in pilot labs
MS2.1.7	60	Full medium registration workflow including integrations to StrainRegistry

5.2.2 M2.2 - Data, Metadata & Workflow Standards

Contributors: EMBL (lead): 50.0 PM, FSU (co-lead): 10.0 PM and in-kind, DSMZ: 6.0 PM and in-kind, JLU: in-kind, RWTH: 10.0 PM, UFZ: 20.0 PM and in-kind, UNIBI: 6.0 PM

Goals: To define, promote, and support the adoption of community-driven, interoperable standards for data, metadata, and analysis workflows in microbiology, ensuring full alignment with the FAIR principles and technical interoperability across disciplines and platforms.

Major accomplishments of first funding period: We primarily focused on aggregating and harmonising technical, environmental, and biological metadata to establish robust metadata frameworks. This work led to the release of a dedicated GitHub repository that consolidates existing, community-adopted metadata standards relevant to microbiological research. The repository serves as both a resource for the scientific

community and the cornerstone of a forthcoming manuscript that will assess the current landscape of data availability and deposition practices. In parallel, virus-related metadata schemas (e.g., GSC MIGSVi and MIGSUViG, ENA Sample Checklists) were collected and mapped to the integrated VirJenDB dataset (M3.3 - Databases & Data Services).

Planned activities for the second funding period: We will shift from compiling metadata standards to actively promoting FAIR-aligned microbiological data submissions, extending our remit from defining standards to ensuring practical usability, sustainability, and semantic interoperability across platforms. Secondly, we will develop and disseminate reproducible, containerised, and community-endorsed workflow standards (e.g., NextFlow, Snakemake; see M3.2 - Analytical Services & Workflows, M4.3 - Workflow Execution Platform) to support transparency and interoperability throughout the data lifecycle. These workflows will streamline processing and enable standardised submissions to public repositories. Thirdly, we will focus on improving microbiology-specific data and metadata submissions to repositories like ENA, which currently lack key microbiology-specific metadata (e.g., host and environmental descriptors). By integrating enriched metadata with standardised workflows, NFDI4Microbiota aims to enhance data quality, interoperability, and reuse M2.5 - Repository Integration & Data Sharing. To address these challenges, the second phase will implement a three-pronged strategy: 1. Development of a microbiology-specific metadata schema: In this task, we will develop customised metadata schemata to capture key domain parameters such as host taxonomy, physiology, health status and fine-grained environmental context through established ontologies (e.g., ENVO, OBI, NCBITaxon). We will document sampling, extraction, sequencing protocols, and analytic workflows to ensure reproducibility, while systematically recording experimental design elements such as treatment groups and time points. For virus data, we will work with VirJenDB (M3.3 - Databases & Data Services) and the GSC to draft the required schemas and extensions. By integrating controlled vocabularies and ontological frameworks, we will enable semantic interoperability and make metadata machine-actionable. This domain-specific approach will allow us to construct high-quality, ontology-driven knowledge graphs in the complex field of microbiome research (M2.4 - Ontology-driven Knowledge Integration). 2. Metadata Standardisation and Submission: To ensure that metadata is machine-actionable and semantically consistent, we will create mapping files aligned with controlled vocabularies and ontologies (e.g., EDAM, ChEBI, BioPortal, OLS). This will eliminate ambiguity from free-text entries and support automated querying, data integration, and downstream analysis – especially for complex health-related or multi-omics data. Comprehensive documentation and clear guidelines will help data providers choose the right terms and remain semantically compliant, lowering the barrier to high-quality metadata submission and promoting interoperability across the microbiome research community. 3. Sustainability and Incentivisation of Metadata Practices: To drive adoption and ensure long-term impact, we will embed sustainability mechanisms from the outset. We will advance high-quality metadata recording practices to improve transparency, reproducibility, and long-term research utility. The metadata-rich submissions will serve as benchmarks for exemplary data stewardship, with DOIs assigned to structured metadata bundles for proper attribution and citation. In

collaboration with M1.3 - Community Engagement, we will develop modular, domain-specific training resources to meet diverse user needs. We will also host inclusive workshops and consultations to co-develop and refine metadata practices, ensuring they are practical, adaptable, and aligned with real-world research needs. This effort directly supports M2.5 - Repository Integration & Data Sharing which focuses on supporting data-generating facilities in submitting well-annotated datasets and embedding FAIR practices at the point of data generation through programmatic tools like Data Producer APIs.

Table 19 lists the milestones for M2.2.

Table 19.
Milestones to be achieved in measure 2.2 - Data, Metadata & Workflow Standards

Milestone	Month	Description
MS2.2.1	1	Expert working groups convened for each data type
MS2.2.2	18	Metadata schema co-designed for all microbiome data types
MS2.2.3	18	Standardised ontologies and customised controlled vocabularies integrated
MS2.2.4	36	Pilot metadata annotation sprints
MS2.2.5	36	Enforcement and automation of metadata templates
MS2.2.6	60	User documentation and guidelines for submission
MS2.2.7	60	Training and outreach

5.2.3 M2.3 - Best Practices for Data Processing

Contributors: HZI (lead): 150.0 PM and in-kind, RWTH (co-lead): 43.0 PM, DSMZ: in-kind, JLU: 8.0 PM

Goals: The overarching goal is to establish and maintain best-practice recommendations for data processing through comprehensive benchmarking of computational methods on complex, realistic datasets. By continuously advancing the CAMI Benchmarking Portal and benchmarking resources, we will help the microbiology community identify reliable, high-performance tools for metagenomics and related analyses, thereby improving the reproducibility, comparability and overall quality of microbiome research.

Major accomplishments of first funding period: Computational metagenomics spans a wide variety of methods, and fast-evolving sequencing technologies now generate ever-larger datasets. To help researchers choose the most suitable software, comprehensive, continuous benchmarking – evaluating both accuracy and efficiency – is essential. Meaningful comparisons require appropriate resources and adherence to community-wide principles such as standardised benchmark datasets, evaluation procedures, established performance criteria, consistent data formats, and the application of FAIR principles. In the first funding period, Measure 2.3 focused on establishing and

continually refining best-practice recommendations for metagenomics data analysis through systematic benchmarking of computational tools. We collaborated with international partners to publish a perspective article in Nature Microbiology on best practices, adoption challenges, and data-sharing opportunities in microbiome research (Huttenhower et al. 2023). We also developed and launched the CAMI Benchmarking Portal (Meyer et al. 2025), released in mid-2024 and already attracting about 15,000 unique users each month. This platform was created in collaboration with the community-driven Critical Assessment of Metagenome Interpretation (CAMI) initiative and builds upon CAMI's extensive benchmarking resources (Meyer et al. 2022). The portal integrates established evaluation tools (MetaQUAST, AMBER, and OPAL) to enable comprehensive assessment, ranking, and visualisation of results for metagenome assembly, genome and taxonomic binning, and taxonomic profiling. Using standardised metrics and CAMI datasets, it supports comparative evaluation of software and pinpoints the most appropriate solutions for specific tasks. The resulting best-practice recommendations were integrated into nf-core based computational workflows and made available on the CloWM platform (M3.2 - Analytical Services & Workflows), enabling users to select appropriate methods with confidence and transparency regardless of theirbioinformatics background.

Planned activities for the second funding period: We will greatly extend the CAMI Benchmarking Portal, in alignment with the recommendations by the NFDI Expert Committee (NFDI Expert Committee 2023, Task c)). A primary goal is to incorporate further data types, particularly metatranscriptomics, which introduces new analytical challenges such as transcript expression quantification, taxonomic redundancy, and differential expression. To support comprehensive benchmarking, we will generate new datasets that reflect a broad diversity of microbial communities, including fungi and viruses, as well as metagenomics data linked to health-related personalised genome collections from isolate sequencing. Many of these datasets will feature novel taxa not yet publicly available, making them valuable for simulating complex, real-world analysis scenarios. In parallel, we will broaden the range of benchmarked method categories to includeapplications such as Biosynthetic Gene Cluster (BGC) detection and pathogen identification. This will enable comparative evaluation of emerging methods and support the development of performance-optimised, application-specific workflows. The benchmarking results will continue to inform best-practice recommendations, which will be communicated through community channels (M1.3 - Community Engagement), training activities (M1.1 - Training & Education), and publications to ensure broad dissemination and adoption in line with the NFDI Expert Committee's recommendation (NFDI Expert Committee 2023, Task b)).

Deliverables:

- Extended Benchmarking Portal: support for new data modalities (e.g. metatranscriptomics) and software categories (e.g. BGC, pathogen detection).
- Creation and release of complex, taxonomically diverse datasets including novel isolates.
- Best-practice guidelines for software and workflow selection.

Table 20 lists the milestones for M2.3.

Table 20.
Milestones to be achieved in measure 2.3 - Best Practices for Data Processing

Milestone	Month	Description
MS2.3.1	24	Biannual novel metagenome benchmarking datasets developed
MS2.3.2	24	Biannual best-practice data-processing recommendation releases for common metagenome analyses
MS2.3.3	48	Biannual novel metagenome benchmarking datasets developed
MS2.3.4	48	Biannual best-practice data-processing recommendation releases for common metagenome analyses
MS2.3.5	52	Metatranscriptome datasets and analytics capabilities developed
MS2.3.6	60	BGC analytics developed

5.2.4 M2.4 - Ontology-driven Knowledge Integration

Contributors: DSMZ (lead): 60.0 PM, ZB MED (co-lead): 30.0 PM and in-kind, FSU: 5.0 PM, RWTH: in-kind

Goals: We will lay the groundwork for comprehensive semantic enrichment of microbiology data, information, and knowledge by curating and applying relevant ontologies to annotate research outputs. This process will standardise the way microbiological data is structured and interpreted, streamline data integration and sharing across the community, and ultimately foster a more interconnected and insightful understanding of microbial species.

Major accomplishments of first funding period: Within the MicrobioKGs Use Case we thoroughly assessed the potential and practical value of knowledge graphs across various microbiology research communities, exploring their role in data integration and discovery and demonstrating their use in real-world applications (Schober et al. 2025).

Planned activities for the second funding period: We will work closely with teams dedicated to data, metadata, and workflow standards (M2.2 - Data, Metadata & Workflow Standards) to ensure that the standards we develop comply with or contribute to existing ontologies. This collaboration will include the creation of a Strain Ontology based on Bac *Dive* and StrainInfo (M3.3 - Databases & Data Services), as well as the development of vocabularies for cultivation media and devices (M2.1 - Experimental Practice & Procedure Standards). Terminology coherence will be maintained across the consortium, in part by reusing established ontologies such as ChEBI, BTO, and ENVO. In parallel, we will work closely with VirJenDB to develop ontologies for virus data based on existing resources (M3.3 - Databases & Data Services). This will help standardise language across our services and strengthen data integration. Additionally, we will datasets into

Resource Description Framework (RDF) using a unified data model, enabling federated queries and seamless integration. In cases where in-house conversion is not feasible, we will empower data providers to perform the conversion themselves. Shared vocabularies and graph-based integration will ensure interoperability among NFDI4Microbiota services, turning fragmented resources into a coherent, user-centric research ecosystem. The resulting graphs will be integrated into the KGI4NFDI infrastructure and published in the central Knowledge Graph Repository, making the data broadly accessible. A unified query interface will be implemented in the NFDI4Microbiota web portal to centrally query all knowledge graphs, which will enhance the user experience and facilitate more-efficient data retrieval. Moreover, we will develop training materials and host workshops for the community, covering topics such as the SPARQL query language to retrieve data and ontology modelling. This will ensure the uptake and reuse of the graphs, fostering a more knowledgeable and skilled research community (M1.1 - Training & Education). In parallel, researchers will be guided to record experimental designs and metadata in electronic lab notebooks (M3.5 - Electronic Lab Notebook) using machine-actionable, knowledge-graph-compliant vocabulary. This guidance will span in vitro, in vivo, and in silico study designs, ensuring research methodologies are consistent and interoperable. Looking ahead, we envision several impactful developments that will extend the reach and utility of an overarching, integrated knowledge graph. By aligning shared ontologies and implementing federated access, we aim to enable seamless cross-domain queries across different NFDI consortia, such as linking microbiota research with biodiversity data from NFDI4Biodiversity. Building on this infrastructure, we envision developing a personalised knowledge assistant that will guide researchers to relevant protocols, datasets, microbial strains, and services tailored to their specific needs. Moreover, the graph will facilitate automated workflows; it can assist in pre-filling metadata forms, recommend suitable SOPs, or enable context-aware data submissions. These developments follow the “FAIR by design” principle, ensuring that data and services are inherently findable, accessible, interoperable, and reusable from the outset.

Deliverables:

- Strain Ontology, media and device vocabularies developed and made publicly available
- RDF conversion toolkit and documentation for data providers
- Knowledge graph integration into KGI4NFDI and joint query interface in NFDI4Microbiota portal
- Training materials and SPARQL workshop series for the community
- Ontology alignment report and shared vocabulary registry

Table 21 lists the milestones for M2.4.

5.2.5 M2.5 - Repository Integration & Data Sharing

Contributors: ZB MED (lead): 20.0 PM, FSU (co-lead): 5.0 PM and in-kind, DSMZ: in-kind, EMBL: 8.0 PM and in-kind, JLU: in-kind, RWTH: 8.5 PM, UFZ: 10.0 PM, UMR: 20.0 PM

Table 21.
Milestones to be achieved in measure 2.4 - Ontology-driven Knowledge Integration

Milestone	Month	Description
MS2.4.1	12	First version of Strain Ontology and cultivation media vocabularies released and published for feedback
MS2.4.2	24	RDF conversion toolkit released and used by at least two data providers for pilot integrations
MS2.4.3	36	Integrated knowledge graphs in collaboration with KGI4NFDI; federated query interface prototyped in the NFDI4Microbiota portal
MS2.4.4	48	Community training program completed with SPARQL and ontology modelling workshops
MS2.4.5	60	Final interoperability report published and cross-consortium querying with NFDI4Biodiversity successfully demonstrated

Goals: To streamline the data flow from data-producing facilities, such as sequencing and proteomics labs, to community-trusted repositories, and to facilitate early, standards-compliant data sharing in line with the “Do It Right The First Time” (DRIFT) concept.

Major accomplishments of first funding period: We piloted direct data brokerage from partner sequencing centres and validated it through a dedicated Flex Fund project at HZI. Several pilot or working submission tools were developed (e.g., the VirJenDB pilot tool for virus data, the ENA submission wizard, and the Qiita instance for amplicon data), some of which were released (e.g., subMG for metagenomics data (Tubbesing et al. 2025)).

Planned activities for the second funding period: We will promote the adoption of tools and practices that enable the early capture of essential provenance, metadata, and contextual information throughout the research process. This early integration is crucial to ensure that microbiological data adheres to the FAIR principles. As part of this effort, we will design metadata templates that are aligned with established community standards such as the ENA checklists and Minimum Information about any (x) Sequence (MIxS) guidelines. These templates will help researchers structure and standardize metadata from the outset. To further support metadata quality, we will implement manual or semi-automated validation steps that allow users to check for completeness, consistency, and compliance before data submission. The tools supporting this metadata compilation will generate identifiers which can be shared with data-generating facilities when samples are submitted, ensuring a clear and traceable link between physical samples and their associated digital metadata. To streamline the flow from data generation to publication, we will establish a data brokerage workflow in which data-generating facilities can directly submit sequencing or measurement data, while researchers provide the corresponding metadata through Aruna (see Services provided by the consortium). Aruna will then combine both components, store them in a structured and versioned format, and enable downstream integration. From there, the data and metadata can be forwarded for further analysis in platforms such as CloWM (see Services provided by the

consortium), or submitted to community-endorsed public repositories like ENA for publication and long-term preservation. NFDI4Microbiota is uniquely positioned to bridge the gap between microbiology researchers, data producers, and international data infrastructures. As the different data-generating facilities have different workflow and operation model, the provided solution has to be rolled out individually. The solution needs to be deployed individually for each data-generating facility due to their unique workflows and operational models.

Deliverables:

- Toolkit for early metadata and provenance capture
- Aruna submission workflow
- ENA submission workflow
- Roll-out plan considering the landscape of the data-generating facility
- Integration guidelines for selected community repositories
- Outreach concept and engagement plan for data-producing labs

Table 22 lists the milestones for M2.5.

Table 22.
Milestones to be achieved in measure 2.5 - Repository Integration & Data Sharing

Milestone	Month	Description
MS2.5.1	9	Tools generated and tested
MS2.5.2	12	Outreach campaign launched to engage pilot data-producing labs M1.3 - Community Engagement & M1.6 - Case Studies)
MS2.5.3	18	Integration of first three data generator facilities completed
MS2.5.4	30	Integration of ten additional data generator facilities completed
MS2.5.5	45	Second toolkit version released based on integration of facilities
MS2.5.6	55	Meetup of integrated facilities conducted

5.2.6 M2.6 - Political Consulting & Legal Frameworks

Contributors: DSMZ (lead): 30.0 PM, EMBL (co-lead): 8.0 PM and in-kind, FSU: 2.0 PM and in-kind, JLU: in-kind

Goals: The aim of this measure is to preserve open, interoperable access to microbial genetic and other biological data by actively shaping United Nations (UN) level policy debates on Digitial Sequence Information (DSI) and providing researchers with guidance and support to meet biodiversity and privacy obligations. It addresses the increasing importance of legal and political frameworks affecting microbiological research.

Major accomplishments of first funding period: While this measure will be newly introduced, some activities on these issues were organised in the first funding period, including an NFDI InfraTalk on the topic of “Current developments in the global Genome Commons and challenges” (NFDI 2022). Community feedback and rising legal concerns have prompted the introduction of a new measure for this task in the second funding period.

Planned activities for the second funding period: We will actively shape international policy discussions by supplying scientific evidence on the implications of emerging DSI obligations within the Convention on Biological Diversity (CBD), the World Health Organisation (WHO) Pandemic Agreement, and the new Biodiversity Beyond National Jurisdictions (BBNJ) Treaty. Policy proposals that threaten open microbial-data governance under the CBD or impose onerous tracking-and-tracing regimes under the WHO and BBNJ, will be countered with empirical data and real-world examples from the NFDI4Microbiota consortium and other NFDI partners, drawing on our MOU with NFDI4Biodiversity and DataPlant (NFDI4Biodiversity 2026). This measure will help researchers meet their legal duties under the CBD and the Nagoya Protocol, raise community awareness, and enable legal compliance without impeding scientific collaboration. Support will be offered through a dedicated Helpdesk, presentations at national and international conferences, and concise legal guidance tailored to microbial data producers. In addition, this measure will help researchers navigate data-protection obligations for human-sensitive data, for example in human microbiome research, particularly in the context of privacy regulations such as the GDPR. We will provide targeted procedural guidance for managing such data in line with national and European standards, closely aligned with technical developments in M3.7 - Privacy Sensitive Data Processing & Secure Processing Environments designed to ensure secure processing on our platforms. These efforts will be coordinated with other infrastructures, including the de.NBI-Cloud, and with other NFDI consortia such as GHGA to ensure interoperability and reusability. Working closely with GHGA, we will harmonise legal, ethical, and technical challenges for sharing sensitive microbiome data. Finally, we will develop tailored user guidelines for responsible data sharing and usage and disseminate them through our services (M3.3 - Databases & Data Services).

Deliverables:

- Position paper or community statement on DSI-related regulations
- Legal guidance brief for microbial researchers on CBD/Nagoya compliance
- Helpdesk integration for legal consultation requests
- Record of community engagement activities (talks, panels, workshops)
- Alignment with GHGA on key legal, ethical, and technical challenges involving human-sensitive microbiome data
- Development of practical guidelines for data governance related to consent, access control, data usage conditions, and data sharing

Table 23 lists the milestones for M2.6.

Table 23.
Milestones to be achieved in measure 2.6 - Political Consulting & Legal Frameworks

Milestone	Month	Description
MS2.6.1	6	Publication of brief legal guidance document on CBD and Nagoya compliance for microbial researchers
MS2.6.2	12	Position statement or expert commentary submitted to DSI policy consultations
MS2.6.3	24	Legal support track implemented in the Helpdesk system
MS2.6.4	36	BBNJ implementation guideline for microbial data management
MS2.6.5	48	WHO Pathogen Access and Benefit Sharing (PABS) implementation guideline for microbial data management
MS2.6.6	58	Summary report on outreach activities and community feedback
MS2.6.7	58	Practical user-centric guidelines for Use Cases around human-sensitive microbiome data

5.2.7 Overall schedule, contribution, risks and dependencies

Fig. 11 illustrates the overall schedule for TA2.

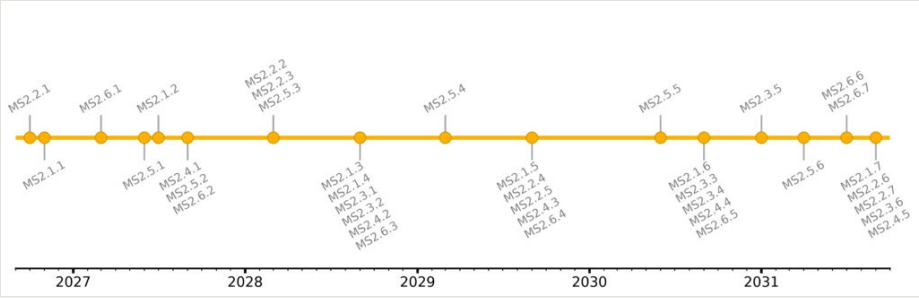

Figure 11.
Overall schedule Task Area 2

Contributions: This TA, which lays important groundwork for many of the consortium's tasks, will be contributed to by eight out of ten co-applicants. Major leadership roles will include: DSMZ for the TA lead, experimental procedure standards, ontologies, data integration, and political and legal consulting; EMBL for data and workflow standards; HZI for data-processing standards; and ZB MED for repository submission. RWTH and FSU will take on measure co-leaderships, while UFZ and JLU will provide in-kind contributions.

Dependencies: Task area 2 - From Bench to Repository: Data Standards & Integration works closely with Task area 1 - Community, Networking & Training to ensure standards are effectively communicated with the community and sufficient training is provided for

the transition to FAIR-compliant repositories. Close collaboration with Task area 3 - Services is needed to embed standards in workflows and services and with Task area 4 - Technical Infrastructure to implement validation and integration mechanisms. Continued alignment with other NFDI consortia, the NFDI section on (meta)data standards, and international initiatives such as ELIXIR, GSC, and International Nucleotide Sequence Database Collaboration (INSDC) is essential for long-term interoperability and adoption.

Risks of implementation and mitigation approaches: Risks include low adoption of standards due to usability barriers or a lack of incentives, and difficulties in accommodating heterogeneous community needs. These will be mitigated by involving user communities early through TA1, co-developing templates with researchers, and iteratively refining standards based on real-world feedback. Technical risks – such as tool incompatibility or schema drift – will be addressed through automated validation services and strong collaboration with Task area 3 - Services and Task area 4 - Technical Infrastructure. The risk of redundancy or fragmentation across NFDI consortia will be mitigated by our active participation in cross-consortium working groups.

5.3 Task area 3 - Services

TA-Coordinators: EMBL (lead), FSU Jena (co-lead)

TA3 aims to provide the microbiology community with integrated services for data access and analysis, lowering technical barriers to standardised and reproducible research (Fig. 12). During the first funding phase, we successfully established the NFDI4Microbiota Hub, a central web portal that connects users to a comprehensive set of high-quality analytical workflows, data management tools, and scientific databases. This includes standardised, FAIR-compliant pipelines covering multiple microbiome-related omics disciplines that are accessible via the CloWM (M4.3 - Workflow Execution Platform), services such as Qiita for comprehensive management of amplicon sequencing data, and essential databases such as VirJenDB and StrainInfo, which provide quality-controlled viral sequence data and reliable strain identifiers, ensuring the sustainability and accessibility of microbial knowledge. Building on this foundation, the second funding phase will create an interconnected ecosystem of analytical and data services, greatly extending the consortium's capabilities and impact. We will boost interoperability across our platform and databases through comprehensive enhancements to web (M3.1 - Central Web Portal) accessibility and mobile support, development of a unified API for seamless service queries, and deployment of a linking hub to enhance the integration and exploration of microbiological data. Integration of our infrastructure with Knowledge Graphs and Model Context Protocols (MCP) will ensure that our data ecosystem is AI-ready. In addition to the development and maintenance of best-practice analytical workflows and services (M3.2 - Analytical Services & Workflows), we will proactively encourage and support the integration of external services from the microbiology community through targeted Flex Fund calls (M5.3 - Dynamic Adaption and Growth). Our expanding toolset will address urgent scientific and societal challenges, such as planetary health and antimicrobial resistance, and support the transition from associative

to mechanistic microbiological research. We will strengthen the field's data foundations (M3.3 - Databases & Data Services) by maintaining and expanding our current databases and safeguarding at-risk resources, thereby improving and securing long-term access to microbial knowledge (Poveda et al. 2025). This includes new databases and workflows for fungi and other microbial eukaryotes, reflecting a growing need in the community. Together with Task area 2 - From Bench to Repository: Data Standards & Integration, we will work to expand and improve the quality and types of metadata submissions to microbiome-related repositories. Embracing the opportunity to act as a data broker for our community, we will support and streamline data deposition by offering services (M3.2 - Analytical Services & Workflows, M3.3 - Databases & Data Services) that simplify the registration of new isolates in strain databases, and that streamline the submission of OMICs data to ENA. While this will contribute to high-quality metadata for future data submissions, we will also FAIR-ify a large volume of existing metagenomics data in ENA via a manually curated metadata catalogue. NFDI4Microbiota will continue to work on the broad adoption of ELNs and will champion open-source solutions (M3.5 - Electronic Lab Notebook). To embed FAIR, reproducible workflows directly into lab practice, we will cultivate an open-source ELN ecosystem, publishing starter templates, funding community-driven plugins, and offering focused training. We will continue working on the long-term preservation of isolate-related metadata (M3.4 - Data Quality & Provenance Services), but also aim to develop and implement an AI-driven metadata enrichment infrastructure for microbial omics entries, mining and structuring metadata from microbiological literature (M3.6 - AI-based Metadata Imputation). As person-specific information becomes more prevalent in microbiome research, the consortium will devise solutions (M3.7 - Privacy Sensitive Data Processing & Secure Processing Environments) in consultation with relevant NFDIs (NFDI4Health and GHGHA) that ensure compliance with the GDPR and national regulations governing personal and sensitive data. Together, these strategic developments will facilitate a robust, interconnected, and accessible infrastructure capable of supporting rapid innovation and community-driven microbiological research in the FAIR and AI-ready era.

5.3.1 M3.1 - Central Web Portal

Contributors: ZB MED (lead): 60.0 PM, EMBL (co-lead): 60.0 PM, DSMZ: in-kind, JLU: in-kind

Goals: To deliver a broad and diverse portfolio of services to the microbiology community in Germany and beyond, all accessible through a single point of entry: the NFDI4Microbiota Hub, a central web-based portal.

Major accomplishments of first funding period: We established the NFDI4Microbiota Hub as a one-stop-shop that connects users to our service portfolio and offers unified search, an extensive knowledge base, a training catalogue, a newsroom, and a dedicated Helpdesk for community support.

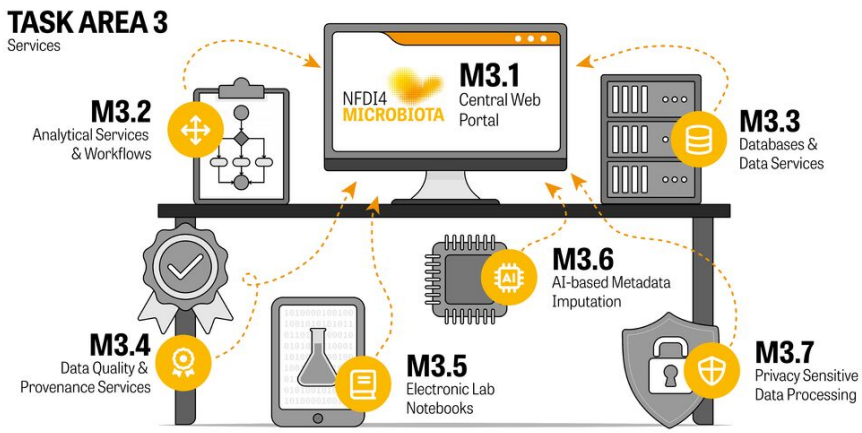

Figure 12.

Overview of Task Area 3

Planned activities for the second funding period: Continuous integration of new content and services will ensure the Hub remains up to date, comprehensive, and responsive to emerging needs. Regular maintenance and the addition of new resources will further expand the platform's capabilities. Accessibility will be improved through barrier-free design, plain language, and mobile-friendly features, while a website overhaul will modernise the visual identity and enhance the overall user experience. A major technical focus will be the development of an overarching API that provides programmatic access to all services, enabling seamless integration with external tools and platforms. The MCP will be implemented to facilitate advanced interoperability, including access by LLMs. Expansion of existing databases and the integration of knowledge graphs will strengthen links across resources, supporting deeper discovery and new data-driven applications within the Hub. New content will be continuously integrated into the platform to ensure that services and resources remain current, comprehensive, and relevant to the community. This ongoing process will facilitate timely updates, incorporation of new data types, and alignment with emerging scientific developments.

Deliverables:

- Continuous integration of content
- Maintenance and inclusion of new services
- Improvements to accessibility
- Rebranding and website overhaul
- Mobile view improvement
- Search functionality improvements
- Overarching API for all services
- Subportal development and database expansion
- MCP integration

- Integration of Knowledge Graphs

Table 24 lists the milestones for M3.1.

Table 24.
Milestones to be achieved in measure 3.1 - Central Web Portal

Milestone	Month	Description
MS3.1.1	3	Features implemented for improved accessibility, including keyboard navigation, text-to-speech, and simplified English for broader accessibility
MS3.1.2	6	Focus on improving mobile optimisation, ensuring responsive design, and enhancing usability on smaller devices
MS3.1.3	12	Search functionality revamped
MS3.1.4	17	Steps initiated to build a subportal to host additional datasets, such as more eukaryotic data, and to improve database integration
MS3.1.5	23	Unified API to query all services developed and implemented
MS3.1.6	30	Knowledge Graphs integrated
MS3.1.7	36	MCP integrated to improve the system's understanding and enhance response capabilities
MS3.1.8	48	Effectiveness of improvements evaluated on a continuous basis, including necessary adjustments to platform services, accessibility features, and search functionality
MS3.1.9	60	Website updated with new design to align with potential new visual identity
MS3.1.10	60	Process established for regular maintenance and integration of new services, ensuring the platform remains up to date and relevant

5.3.2 M3.2 - Analytical Services & Workflows

Contributors: EMBL (lead): 90.0 PM and in-kind, UNIBI (co-lead): 27.0 PM and in-kind, DSMZ: in-kind, FSU: 10.0 PM and in-kind, JLU: 46.0 PM, RWTH: 23.0 PM

Goals: This measure develops, maintains, and continually integrates best-practice analytical workflows and services, creating a sustainable, visible, and user-friendly suite of tools that enables microbiology researchers to process microbiology-related data in a FAIR and reproducible manner.

Major accomplishments of first funding period: During the first funding period, we released an extended version of the nf-core/mag workflow, incorporating best-practice tools identified in the second CAMI challenge (M2.3 - Best Practices for Data Processing) to deliver a fully benchmarked pipeline from raw reads to high-quality Metagenome-Assembled Genomes (MAGs). We also deployed a public Qiita instance for researchers to manage amplicon studies entirely in the browser, from metadata capture through to final analyses. Consortium partners contributed a further 30 production-ready workflows

covering various OMICs disciplines, all compliant with the consortium's standards (M2.2 - Data, Metadata & Workflow Standards) and available via open-source repositories and on NFDI4Microbiota's workflow execution platform, the CloWM (M4.3 - Workflow Execution Platform). Consequently, NFDI4Microbiota's analytical services currently cover amplicon sequencing, bacterial and archaeal genomics, short- and long-read metagenomics, transcriptomics, epigenomics, and integrative multi-omics pipelines, alongside specialised tools for pangenome graph construction, strain tracking, phylogenomics, and viral-genome assembly.

Planned activities for the second funding period: Building on this foundation, NFDI4Microbiota will accelerate the onboarding of high-quality workflows and services, tightly co-developed with M4.3 - Workflow Execution Platform, to reinforce CloWM and increase pipeline robustness. A new process, linked to the consortium's Flex Fund scheme, will actively recruit the wider community to fill gaps in our portfolio, supporting the integration of external contributions to meet evolving needs while also drawing on our collaboration with projects such as nf-core and the Galaxy community. In particular, we will enrich our catalogue with newly recruited and developed workflows and services that address emerging scientific and societal challenges; we have already identified four key topics that promise the greatest benefit for users and for the wider microbiome field: 1.) Planetary health: Accelerating climate change and the spread of Antimicrobial Resistance (AMR) are converging threats to our planet and societies, reshaping ecosystems and challenging healthcare systems worldwide. To enable the systematic study of planetary health, we will develop and release workflows for annotating mobile genetic elements, detecting horizontal gene transfer events, phages, and profiling AMR genes with input from the VirJenDB (M3.3 - Databases & Data Services). We will further enhance the Metagenomics-Toolkit (Belmann et al. 2025) workflow to allow for the recovery of plasmids, unassembled microbial community members, and the discovery of microbial inter-dependencies in these ecosystems. We will also leverage our partners' involvement in the TREC expedition, which generated paired multi-omics data along the European coastline. By integrating and publishing the multi-OMIC TREC workflows to CloWM, we will connect the community to this flagship dataset and enable them to chart ecosystem responses from molecules to communities. Fungi and other microbial eukaryotes are essential, yet understudied components of planetary biodiversity. To foster insights into this domain, we will develop workflows for profiling, assembling, and annotating eukaryotic genomes, enable their interactive analysis through Qiita, and provide standardised submission to ENA M2.5 - Repository Integration & Data Sharing together with the registration of high-quality strains via MicrobeDive (M3.3 - Databases & Data Services). Since these workflows are highly compute-intensive, their deployment on our elastic cloud infrastructure is ideally suited to benefit our community. 2.) From association to mechanism: The microbiome field is currently advancing from correlative research towards functional work that can lead to mechanistic insights, as exemplified by the newly funded Priority Program SPP2474 "Illuminating Gene Functions in the Human Gut Microbiome". We will support these efforts by intensifying our activities at two levels. First, to help researchers comprehensively capture functional potential and activity at the community level, we will critically benchmark reference-free de-novo metatranscriptomic

assemblers in M2.3 - Best Practices for Data Processing and make the resulting best-practice workflows available on CloWM. Second, we will support promising developments in isolate-centric high-throughput omics platforms, providing workflows and services (in collaboration with M3.3 - Databases & Data Services) that will enable laboratories to rapidly progress from cultivation to FAIR-compliant genomes and phenotypes (Clavel et al. 2025). Examples here include dereplication of mass spectrometry data (Pauvert 2023, Pauvert 2025) and their analysis for taxonomy and function, including genus-level classification. 3.) Interconnected service ecosystem: We will support the microbiological research community in developing new microbiome-related services and integrating them into a vibrant, interconnected ecosystem accessible through the central web portal (M3.1 - Central Web Portal). First, we will connect partners' existing services (e.g., MetaTraits and VirJenDB) via the CloWM API, documenting and showcasing its utility as a robust, scalable backend that allows specialised front-end applications to delegate compute-intensive tasks to a shared cloud infrastructure (M4.1 - Computational Infrastructure Operations and Software Components). Second, we will deploy SPIRE view, an interactive viewer for taxonomic profiles generated by widely used amplicon and metagenomic tools (MetaPhlAn, mOTUs, Kraken2/Bracken, Krakenuniq, Kaiju, QIIME 2). SPIRE view will visualise microbial community composition, provide basic quality assessments, and enable taxonomy translation between National Center for Biotechnology Information (NCBI) and Genome Taxonomy Database (GTDB). Importantly, it will function as a central linking hub, offering direct, one-click access from each taxon to complementary resources such as trait annotations (MetaTraits), strain metadata (Bac*Dive*), or ecological predictions (Enterotyper, MicrobialLoadPredictor), among others. 4.) Bench-to-Repository: Realising our bench-to-repository vision for amplicon data, we will actively help users meet high quality standards for sequence and metadata submissions to long term repositories (e.g., ENA), brokered directly through Qiita, and we will connect Qiita to our partnering sequencing centres (M2.5 - Repository Integration & Data Sharing). Through additional plugins, we will extend the data types that can be managed in Qiita (e.g., to include fungal isolate annotations).

Deliverables:

- Integration of MetaTraits services with CloWM API
- Launch of workflows for horizontal gene transfer and antimicrobial resistance profiling
- Public release of TREC multi-omics workflows on CloWM
- Deployment of specialised workflows for microbial isolate processing
- Release of eukaryotic microbial genome assembly and annotation workflows
- Metatranscriptome assembler workflows benchmarked and integrated into CloWM
- Enhancement of Metagenomics-Toolkit[1] workflow
- Full release of SPIRE view for interactive taxonomic profiling and integration with linked services

Table 25 lists the milestones for M3.2.

Table 25.
Milestones to be achieved in measure 3.2 - Analytical Services & Workflows

Milestone	Month	Description
MS3.2.1	12	Workflows for mobile genetic elements and antimicrobial resistance profiling
MS3.2.2	24	Public availability of TREC-adapted CloWM workflows
MS3.2.3	24	CloWM-integrated pipelines for microbial isolate genome processing
MS3.2.4	30	CloWM workflows for eukaryotic microbial genome analysis
MS3.2.5	36	Benchmarking report and CloWM integration of metatranscriptomic assembly workflows
MS3.2.6	42	Enhanced Metagenomics-Toolkit workflow released
MS3.2.7	48	SPIRE view fully operational for public use
MS3.2.8	60	Extended Qiita functionalities for high-quality metadata and sequence submission to ENA and fungal isolate annotation

5.3.3 M3.3 - Databases & Data Services

Contributors: DSMZ (lead): 120.0 PM, FSU (co-lead): 54.0 PM, EMBL: 15.0 PM, JLU: in-kind, UFZ: in-kind, ZB MED: in-kind

Goals: This measure aims to improve access to and use of microbial knowledge by providing reliable scientific databases and data services to the research community. This includes developing new resources to fill knowledge gaps, consolidating key databases, and safeguarding data from at-risk sources to prevent knowledge loss and thus addresses recommendations b), c) and d) of the NFDI Expert Committee (NFDI Expert Committee 2023).

Major accomplishments of first funding period: Two new databases, StrainInfo (Lissin et al. 2025) and VirJenDB, were successfully developed and are now in productive use, resolving microbial strain identifiers and providing quality-controlled virus sequence data, respectively. The first StrainInfo.net database (2006–2019) (Van Brabant et al. 2006) was a valuable resource with more than 20 000 users per month, before going offline after losing funding. Within the NFDI4Microbiota consortium, we successfully developed a new StrainInfo database by building a new infrastructure integrating the highly valuable data from the old database as well as newly mobilised data. StrainInfo now resolves 717 337 strain identifiers representing 314 035 microbial strains and links them to culture-collection catalogues, sequence databases, and scientific literature, thereby closing a major identification gap and providing a unique service to make strain-associated data FAIR. By assigning Digital Object Identifiers (DOIs) to strains and providing an API for integration into other services, StrainInfo enables unambiguous, persistent citation of microbes across the research ecosystem. VirJenDB was initiated to address community needs for the standardisation of virus (meta)data and to consolidate and improve

reproducibility of distributed, redundant curation efforts across all viruses. After evaluating sources of virus sequences and metadata (Ritsch et al. 2023), we proceeded to harvest, integrate, and merge a large dataset of virus metadata and sequencing data from primary repositories, secondary databases, and standards. Metadata curation and sequence clustering were performed with fully reproducible workflows available on our GitHub repository in line with task d) of the NFDI Expert Committee (NFDI Expert Committee 2023). The dataset is updated continuously and, as of August 2025, contains 15.4 million virus records with 200 metadata fields, all accessible through the VirJenDB portal. As a new service of the consortium, VirJenDB was developed on the de.NBI cloud (M4.1 - Computational Infrastructure Operations and Software Components) and provided a pilot case for integration with the Aruna system (M4.2 - Data Storage Platform). User feedback from the virus community interest group M1.3 - Community Engagement has steadily improved the portal's usability. In addition, virus workflows were recruited for CloWM from the developer community (M4.3 - Workflow Execution Platform) and expertise in virus resources and bioinformatics was contributed to NFDI4Microbiota's training and outreach activities (M1.1 - Training & Education).

Planned activities for the second funding period: We will extend and enhance the established StrainInfo and VirJenDB services, while launching new resources: a central registry that streamlines microbial-strain deposition, a dedicated repository for eukaryotic microbial strain data, and a database that harmonises metagenome-sequence metadata. Further, we will share our expertise to support promising new projects (M1.6 - Case Studies) and to find solutions for orphaned databases and datasets (Poveda et al. 2025). Database services will engage closely with CIGs M1.3 - Community Engagement to improve feedback collection and training materials (M1.1 - Training & Education), thereby fulfilling task b) of the NFDI Expert Committee (NFDI Expert Committee 2023). All database services will be built using ontologies and data standards consolidated or developed by the NFDI consortia (M2.2 - Data, Metadata & Workflow Standards). Datasets, knowledge graphs (M2.4 - Ontology-driven Knowledge Integration), workflows (M3.2 - Analytical Services & Workflows), and tools developed as part of the database services will be fully integrated into the consortium's service portfolio (M4.4 - Base4NFDI and EOSC Service Integration) and made accessible through the central NFDI4Microbiota web portal (M3.1 - Central Web Portal). Support for database services will be provided through the cross-consortial HelpDesk and other platforms (M1.2 - Helpdesk & Support), while training on these services will be embedded in the consortium's training programme (M1.1 - Training & Education) in line with tasks b) and c) of the NFDI Expert Committee (NFDI Expert Committee 2023). Feedback from community interest groups (M1.3 - Community Engagement) and privacy-sensitive service monitoring (M4.5 - Service Monitoring & Reporting) will drive iterative improvements. Within this framework, four major database services will be developed by NFDI4Microbiota: 1.) StrainInfo and StrainRegistry: We will consolidate StrainInfo into a high-quality, long-term resource for microbial researchers, thereby fulfilling task c) of the NFDI Expert Committee (NFDI Expert Committee 2023). Its data model will be enriched with additional fields such as legal restrictions (M2.6 - Political Consulting & Legal Frameworks), and we will actively promote the use of Strain DOIs in scientific publications

and ELN (M3.5 - Electronic Lab Notebook). Concurrently, we will develop standalone tools such as strain-name matching, entity recognition, and the automated imputation of strain identity, all designed for integration into bioinformatics workflows (M3.6 - AI-based Metadata Imputation). A major new step will be the integration of StrainRegistry—a central strain registry and deposition management system—into StrainInfo, an idea born from community feedback during the first funding period. Large microbiome studies now produce many isolates that must be deposited in culture collections, both to support open science and to formally name any new taxa found among them. Researchers need culture collection numbers as early as possible to use as persistent identifiers, yet collections often lack the capacity for bulk submissions. StrainRegistry will remove this bottleneck by allowing microbiologists to register strains, supply full metadata and deposition details, and submit strains directly to participating collections. Strain identity data will be integrated into StrainInfo, while additional services will support the optional transfer of sequence, phenotypic, or cultivation data to specialised databases (M2.1 - Experimental Practice & Procedure Standards, M2.5 - Repository Integration & Data Sharing). With minimal effort for researchers, this system will deliver persistent referencing, close integration with sequence and literature information, and the creation of FAIR, visible datasets. 2.) VirJenDB: Building upon our previous work, we will continue to provide regular releases of our large, curated, AI-ready dataset. Collaboration with key international virus resources (ViralZone, BV-BRC, NCBI Virus/Taxonomy/ENA/INSDC, ICTV, and DSMZ databases) M1.5 - Connection to International Partners and communities will continue, fulfilling task c) of the NFDI Expert Committee (NFDI Expert Committee 2023). To increase interoperability of virus data across research domains, we will facilitate the development of draft metadata schema extensions in cooperation with the GSC (M2.2 - Data, Metadata & Workflow Standards) and help increase and enrich submissions to ENA/INSDC, with a particular focus on mobilising and reusing German virus sequencing datasets (M2.5 - Repository Integration & Data Sharing). We will continue to use the Aruna data storage system and recruit and embed virus workflows from CloWM in line with task d) of the NFDI Expert Committee (NFDI Expert Committee 2023). Through the VirJenDB portal, we will coordinate community-driven metadata curation and correction efforts for virus data (M3.4 - Data Quality & Provenance Services). 3.) MicrobeDive: Within the StrainInfo database, 54 % of strains (168 919 entries) are eukaryotic, yet the information we gather is limited to the strain identity. While comprehensive data on prokaryotic strains will be integrated into the Global Core Biodata and ELIXIR Core Data Resource Bac*Dive* (Schober et al. 2025), no central repository currently exists for eukaryotic strains. Building on community feedback, we will remedy this by adding a fungal and algal section to Bac*Dive* and launching the overarching MicrobeDive, thereby fulfilling task b) of the NFDI Expert Committee (NFDI Expert Committee 2023) to broaden coverage of user communities. Key to this development will be the creation of a standard for microbial strain data (M2.2 - Data, Metadata & Workflow Standards) which will streamline the provision, exchange, and updating of strain information across different sources. Integration of the nomenclatural databases Mycobank and Species fungorum will ensure up-to-date nomenclature. Cultivation procedures and media recipes will be closely connected to the Media*Dive* database (M2.1 - Experimental Practice & Procedure Standards). 4.) Metalog: Developed by the

Bork Group (EMBL Heidelberg), Metalog aims to make metagenomics metadata FAIR and readily accessible to the scientific research community. As a manually curated repository that meets ENVO and UBERON standards, this resource currently holds metadata for 126 000 metagenomes from 834 global studies across diverse environments. In the second funding period, Metalog will be integrated into NFDI4Microbiota's infrastructure, including the Knowledge Graph and search functionalities, as well as analytical services such as MetaTraits trait profiles. Metadata curation efforts will expand to include publicly available contextual geo-referenced datasets, such as the IUCN Global Ecosystem Typology (Keith et al. 2022), significantly enhancing metadata quality, discoverability, and interoperability. Flex Funds for promising projects and data rescue: To make the most of NFDI4Microbiota's expertise and resources, we will use Flex Funds (M5.3 - Dynamic Adaption & Growth) to create new databases and services and to safeguard at-risk data. These efforts will focus on standardising data using ontologies, providing it through an NFDI4Microbiota knowledge graph integrated into the NFDI4Microbiota webportal, and ensuring it is sustainable, FAIR, and AI-ready. Rapid changes in the global infrastructure landscape compel us to reserve resources so we can respond swiftly to emerging community needs in accordance with task d) of the NFDI Expert Committee (NFDI Expert Committee 2023). To keep database development lean, we will focus on the use of open infrastructures such as Wikidata, Wikibase Cloud and the Aruna platform (M4.2 - Data Storage Platform) and aim for community driven additions that could be supported by Flex Fund. Potential projects include ZINC22 DB, a unique database of 54.9 billion small molecules that powers drug-discovery and microbiology research; MOSAIC, a new database of standardised reusable genetic parts that streamlines modular cloning; and BakRep, a large-scale web repository of bacterial genomes enriched with standardised genome characterisations and related metadata. All candidates will be assessed through the established, open and fair Flex Fund evaluation process.

Deliverables:

- VJDB: Release of embedded CloWM virus bioinformatics workflows
- VJDB: Regular releases of the curated dataset through the portal and API
- Full release of the StrainRegistry for automated submission of new strains
- Full release of MicrobeDive providing integrated taxonomy for fungal data
- Metalog: regular releases of curated Metalog dataset through the API and web interface

Table 26 lists the milestones for M3.3.

5.3.4 M3.4 - Data Quality & Provenance Services

Contributors: UMR (lead): 30.0 PM, FSU (co-lead): 5.0 PM and in-kind, JLU: 6.0 PM, UFZ: 20.0 PM and in-kind

Goals: The goal of Data Quality & Provenance Services is to ensure that bioinformatics data and analyses are trustworthy, transparent, and reproducible. This involves tracking

how data are generated and processed, maintaining clear records of all steps and parameters, and supporting validation through well-documented, high-quality metadata. By promoting consistent standards and practices, these services aim to improve scientific rigour, enable reproducibility, and foster collaboration across the research community.

Table 26.
Milestones to be achieved in measure 3.3 - Databases & Data Services

Milestone	Month	Description
MS3.3.1	6	VJDB: First version of metadata schema extension
MS3.3.2	12	Extension of StrainInfo data fields and functions
MS3.3.3	18	First functional prototype of StrainRegistry providing a curation interface
MS3.3.4	18	VJDB: First user data uploaded to workbench
MS3.3.5	24	First functional prototype of MicrobeDive providing mobilised data
MS3.3.6	24	VJDB: First training demo using the workbench
MS3.3.7	24	Metalog: Functional integration of Metalog metadata into NFDI4Microbiota Search and Knowledge Graph
MS3.3.8	60	Fully integrated data submission workflow for strain data through StrainRegistry into MicrobeDive and MediaRegistry
MS3.3.9	60	VJDB: First scientific publication based on dataset
MS3.3.10	60	Metalog: Continuous metadata curation and integration with GPS-based contextual data

Major accomplishments of first funding period: The Provenance Guidelines for Workflow Developers (Klemm 2024) were successfully released to help both new and experienced developers to create bioinformatics workflows with robust provenance tracking, reproducibility, and transparency; several complementary guidelines on data provenance and quality were also issued. Building on these achievements, the MicroProv project was launched to physically link digital metadata with biological samples.
Developed for wet-lab researchers, MicroProv encodes provenance information—such as DOIs, protocols, Open Researcher and Contributor ID (ORCID)/Research Organization Registry (ROR) IDs, and repository links—into plasmid Deoxyribonucleic Acid (DNA) in a compact, error-resistant format. The embedded metadata can be retrieved via standard Sanger sequencing, enhancing traceability and reproducibility of lab experiments.

Planned activities for the second funding period: In the second funding phase, we will make the solution more accessible by developing an open-source graphical interface to streamline use in lab settings. We also plan to expand MicroProv to tag bacterial strains and cell lines using a CRISPR-based system for genome integration, broadening its applicability and user base. Furthermore, the topic of sustainable access to research

software will be addressed: Digital preservation already protects research data against shifts in technology, standards, and software, ensuring that they remain FAIR and reusable. Yet solutions for long-term reusability of research software and AI models are not widely adopted, even though the microbiology community relies ever more on published software. NFDI4Microbiota will address this gap through its involvement in the NFDI “Common Infrastructures” section and will organise up to two workshops dedicated to research-software and AI model preservation.

Deliverables:

- Ontology for dbBact
- Metadata annotation plugin for Qiita
- MicroProv software suite (Graphical User Interface (GUI)/Command Line Interface (CLI)) for embedding provenance in DNA
- Protocol for DNA-encoded provenance (wet-lab + computational)
- Research-software and and AI model preservation workshops

Table 27 lists the milestones for M3.4.

Table 27.
Milestones to be achieved in measure 3.4 - Data Quality & Provenance Services

Milestone	Month	Description
MS3.4.1	6	Release of MicroProv CLI published on GitHub
MS3.4.2	18	Web platform and database for MicroProv GUI published
MS3.4.3	24	Extension of MicroProv for genomic tagging
MS3.4.4	36	DNA-encoded provenance protocol released via protocols.io
MS3.4.5	60	Interactive metadata enrichment plugin online

5.3.5 M3.5 - Electronic Lab Notebook

Contributors: UMR (lead): 30.0 PM, ZB MED (co-lead): 12.0 PM, DSMZ: in-kind, RWTH: in-kind

Goals: This measure aims to establish a sustainable and extensible ELN ecosystem that supports digital-first research practices in microbiology. This will be achieved by developing targeted plugins, integrating workflows, and enhancing the machine-readability of open-source ELN. Adoption will be further accelerated by providing user-friendly templates for core research workflows.

Major accomplishments of first funding period: In the first funding phase, NFDI4Microbiota conducted a comprehensive survey of members’ use of ELN tools, hosted a World Café to identify community needs (Wilbrandt et al. 2025), and launched an annual ELN seminar series to foster knowledge exchange. We also developed a work

plan for adopting open-source ELNs, including a published implementation strategy (Vandendorpe et al. 2024b). Beyond these planned activities, we partnered with Julius Maximilian University of Würzburg to test open-source ELNs, delivered more than 30 training sessions to around 800 participants (see example slides in Vandendorpe et al. (2024a) and accompanying demo video at ZB MED - Informationszentrum Lebenswissenschaften et al. (2024)), and provided tailored consultations on tool selection. Despite these efforts, adoption remained limited due to several persistent challenges: many commercial ELNs impose vendor lock-in, open-source alternatives sometimes present steep learning curves, and many groups struggle to configure a meaningful starter instance, leaving them with an empty notebook and little guidance. Cross-institutional collaboration was hampered by non-interoperable data formats, and few solutions seamlessly integrated with our existing services (such as DMP4NFDI/ Research Data Management Organizer (RDMO), protocols.io, or Persistent Identifier (PID) registries).

Planned activities for the second funding period: Building on insights from the first funding phase, the next phase will move from generic advice to hands-on, community-driven implementation. We will champion open-source ELNs, particularly eLabFTW, as reference implementations and solicit targeted plugin and API development through a Flex Fund call. This call will support the integration of Data Management Plan workflows (directly or via RDMO links), bi-directional SOP exchange with platforms such as protocols.io, and in-line resolution of persistent identifiers (e.g., DOI, StrainInfo, Bac*Dive*/ Media*Dive*, ENA/Sequence Read Archive (SRA)). Efforts will also advance machine-readability of ELNs to enable automated provenance capture, speech-driven hands-free logging, and AI-assisted experiment planning based on historical entries. To accelerate adoption, user-friendly starter templates for core microbiota workflows will be developed, covering topics such as strain inventory management, instrument booking, large-scale collaborative projects, PhD progress tracking, and teaching labs. These templates will be iteratively improved via targeted workshops and train-the-trainer programs, and published on an open template marketplace. Selected community-created templates will be curated and re-released as NFDI4Microbiota-standardised versions to ensure maximum visibility and broad distribution. Complemented by regular data steward consultation hours (see M1.2 - Helpdesk & Support) and published case studies, this approach will equip research groups with fully configured, FAIR-aligned ELNs that integrate seamlessly into the NFDI4Microbiota ecosystem. Continuous improvement will be guided by a robust feedback loop that uses usage metrics and user interviews to refine templates, integrations, and training materials. Further development will focus on integrating ELNs with Laboratory information management systems (LIMS) to connect experimental planning with sample tracking and analysis pipelines, enabling real-time data visualisation within ELNs, and exploring federated querying across ELN instances for large-scale meta-analyses. Developed in collaboration with consortium partners and based on open standards, these efforts will establish a community-driven, extensible ELN ecosystem that provides a strong foundation for future RDM innovation and supports digital-first lab practice in microbiota research.

Deliverables:

- Announcement and evaluation of Flex Fund call for ELN plugin/API development
- Curated starter template packs for microbiota research workflows published on the ELN marketplace
- Detailed workshop materials and train-the-trainer curricula
- Production-ready integration modules for RDMO/DMP4NFDI, protocols.io, and PID registries
- Published machine-readable ELN schema and functional AI-driven prototype

Table 28 lists the milestones for M3.5.

Table 28.
Milestones to be achieved in measure 3.5 - Electronic Lab Notebook

Milestone	Month	Description
MS3.5.1	3	Flex Funds call for ELN extensions and API plugins launched and promoted
MS3.5.2	9	Integration modules for RDMO/DMP4NFDI and protocols.io delivered
MS3.5.3	12	First starter template pack (strain inventory, teaching labs) released
MS3.5.4	24	Targeted workshops and train-the-trainer sessions conducted with ELN developers
MS3.5.5	36	Machine-readable ELN schema published; AI-driven prototype deployed
MS3.5.6	48	Full rollout of the integrated ELN ecosystem; community feedback collected

5.3.6 M3.6 - AI-based Metadata Imputation

Contributors: HZI (lead): 80.0 PM and in-kind, ZB MED (co-lead): 10.0 PM, DSMZ: 7.2 PM and in-kind

Goals: This measure aims to develop and establish a sustainable, AI-driven infrastructure that enriches microbial data sets by mining the microbiological literature and structuring technical, biological, and environmental metadata across taxonomic ranks.

Major accomplishments of first funding period: This measure will be introduced in the second funding phase. It will expand the ongoing DSMZ-HZI collaboration, build on previously developed models (Koblitz et al. 2025), and leverage the StrainInfo database developed during the first funding period.

Planned activities for the second funding period: We will create an AI-driven pipeline that automatically mines the microbiology literature, extracts well-sourced metadata at the strain and species level for more than 10 000 microbial taxa, and feeds the curated information back into key repositories such as ENA (M2.5 - Repository Integration & Data Sharing). To this end, we will systematically harvest titles, abstracts, and full texts from established biomedical and life science repositories, including PubMed, Europe PMC,

bioRxiv, and related repositories, where licenses allow. Using the identifier services provided by StrainInfo (M3.3 - Databases & Data Services), all organism and strain names will be mapped to their corresponding entries in Bac*Dive*, resulting in a taxon-centric corpus expected to cover 60 % to 70 % of the target strains and species. Based on this corpus, we will apply open retrieval-augmented LLMs such as Llama 3-70B or DeepSeekR1/V3. Through dynamically constructed prompts, these models will navigate the literature to extract information on growth conditions, sequencing technologies, host organisms, and other relevant traits. Each extracted item will be linked to its source via precise citations, evaluated with confidence scores, and flagged for expert review at the DSMZ. Since metadata only provide lasting value when aligned with community standards, an agent-based framework will complement the LLM-based extraction pipeline. This framework will invoke established bioinformatics tools to validate file formats, perform quality checks, and convert outputs into interoperable forms as required. It will directly contribute to the goals outlined in M2.3 - Best Practices for Data Processing, enabling a systematic assessment of current metagenomics software and the formulation of best-practice recommendations for data processing workflows. The curated results will flow into a central knowledge graph (M2.4 - Ontology-driven Knowledge Integration), continuously synchronised with Bac*Dive* and Microbe*Dive* (M3.3 - Databases & Data Services), and made accessible via a SPARQL endpoint to support large-scale structural queries. Additionally, this infrastructure will serve as the foundation for a Retrieval-Augmented Generation-based (RAG) chat system integrated into the NFDI4Microbiota webportal (M3.1 - Central WebPortal) that allows users to query the data in natural language. The final outcome will be a sustainable, evidence-based metadata curation service that reduces researchers' workload while improving the overall quality and reusability of microbial datasets, in line with the strategic aims of the WR (German Science And Humanities Council 2025). The underlying text corpus will be continuously expanded, with newly extracted metadata integrated to ensure long-term relevance and impact for the research community.

Deliverables:

- Curated, taxon-centric literature corpus mapped to StrainInfo/Bac*Dive* identifiers
- Open-source, AI-driven metadata extraction pipeline
- Agent-based validation toolkit yielding FAIR-compliant metadata
- Public API & SPARQL endpoint for the knowledge graph
- RAGchat interface embedded in the NFDI4Microbiota web portal

Table 29 lists the milestones for M3.6.

5.3.7 M3.7 - Privacy Sensitive Data Processing & Secure Processing Environments

Contributors: UMR (lead): 30.0 PM and in-kind, EMBL (co-lead): in-kind, FSU: in-kind, JLU: in-kind

Goals: This measure aims to ensure privacy, security, and regulatory compliance in microbiome research involving sensitive personal and clinical data. To achieve this, we will develop and deploy tools and best practices that support robust anonymisation, secure processing, and compliant sharing of microbiome data and metadata.

Table 29.
Milestones to be achieved in measure 3.6 - AI-based Metadata Imputation

Milestone	Month	Description
MS3.6.1	12	Comprehensive taxon-centric literature corpus assembled and validated by DSMZ experts
MS3.6.2	36	Retrieval-augmented or in-context metadata extraction framework deployed and serving enriched strain data via prototype REST/GraphQL endpoints
MS3.6.3	60	Finalised, continuously synchronised integration with Microbe*Dive*/Bac*Dive* and the central knowledge graph, documented and publicly accessible

Major accomplishments of first funding period: This is a novel aspect within this community. It will address the emerging need for rulesets and best-practice examples of how to handle large sets of privacy-sensitive data in a research environment.

Planned activities for the second funding period: Microbiome research often includes personal information such as health status, age, location, diagnoses, treatment history, and samples from various body sites, all accompanied by detailed metadata. Under GDPR and similar regulations, much of this qualifies as sensitive data and must be handled with strict access controls, secure storage and analysis environments, and, where appropriate, pseudonymisation or full anonymisation for public data release. To meet these requirements, our approach combines technical infrastructure, collaborative workflow development, and targeted training. We will establish SPE using Secure Enclaves (Secure Encrypted Virtualization (SEV)/Secure Memory Encryption (SME)) and containerised solutions (e.g. in the de.NBI Cloud) to enable controlled and isolated handling of sensitive data. Alongside this infrastructure, we will provide training for service providers and end users on implementing secure workflows. We will also develop concepts to enable processing of privacy-sensitive data with our workflow management system CloWM. These workflows will be implemented and tested in close collaboration with the GHGA and NFDI4Health, strengthening interoperability and compliance across systems. GHGA's expertise in managing human omics data will inform our metadata handling strategies, while CloWM provides the technical foundation for secure and reproducible workflow execution. Together with GHGA, we will adopt and adapt its metadata schemas, controlled vocabularies (e.g., Human Phenotype Ontology (HPO), Disease Ontology ID (DOID)), and data-access frameworks for microbiome studies involving human participants. This will ensure that sensitive metadata such as health status, cohort identifiers, and phenotypic variables can be appropriately curated, pseudonymised, and deposited in the controlled-access repositories maintained by GHGA and European Genome-phenome Archive (EGA). We will also explore interoperability models between GHGA and open repositories such as ENA, enabling

dual-track submissions where non-sensitive data and metadata are made openly available, while personal and clinical metadata are securely stored and shared under controlled conditions. This hybrid approach maximises the FAIRness of microbiome data without compromising ethical or legal requirements. In accordance with task h) of the NFDI Expert Committee (NFDI Expert Committee 2023), these collaborative activities are central to our efforts to deepen national and international networking (see M2.6 - Political Consulting & Legal Frameworks) and promote cross-consortia reuse of data and services.

Deliverables:

- Anonymisation tools for patient data in cancer research
- Secure Virtual Environments using SEV/SME and containerised solutions
- Comprehensive data security and privacy strategy for processing sensitive patient data
- Documentation and guidelines for anonymisation and secure processing workflows
- Training material for service providers on how to use anonymisation tools and secure environments
- Secure data-sharing protocols for collaborative research while maintaining patient confidentiality

Table 30 lists the milestones for M3.7.

Table 30.
Milestones to be achieved in measure 3.7 - Privacy Sensitive Data Processing & Secure Processing Environments

Milestone	Month	Description
MS3.7.1	12	First prototype of anonymisation tools for patient metadata developed and tested
MS3.7.2	18	Implementation of Secure Virtual Environments (SEV, SME, containers) for patient data processing
MS3.7.3	24	Secure data-sharing protocol finalised and tested for collaborative research use
MS3.7.4	30	Documentation and guidelines for anonymisation and secure processing workflows published
MS3.7.5	36	Final evaluation and optimisation of privacy-sensitive data processing tools for large-scale research

5.3.8 Overall schedule, contribution, risks and dependencies

Fig. 13 illustrates the overall schedule for TA3.

Contributions: The majority of work in this task area revolves around the provisioning of services for end-users. All co-applicants are involved in at least one of the individual

measures, with major leading roles taken by ZB MED, EMBL, DSMZ, UMR, and HZI. The co-applicants will contribute their diverse technical and domain-specific expertise to the development of user-friendly services.

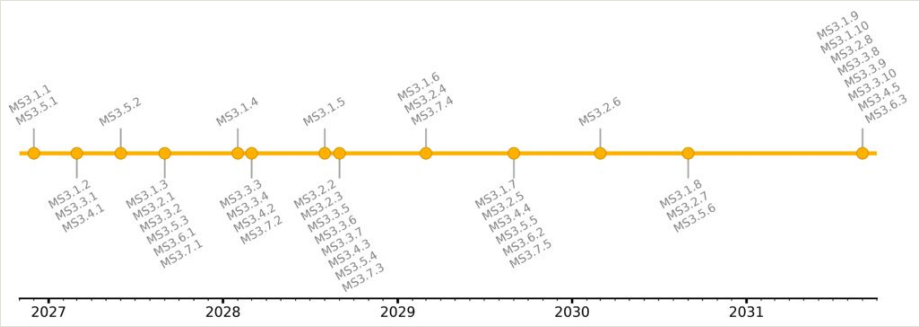

Figure 13.
Overall schedule Task Area 3

Dependencies: The data and analytical services implemented in Task area 3 - Services will utilise the infrastructure provided by Task area 4 - Technical Infrastructure. This includes leveraging the M4.3 - Workflow Execution Platform for deploying Nextflow and Snakemake-based workflows, as well as utilising compute backends such as SLURM or Spark clusters for more advanced, scalable data analysis. The storage platform Aruna (M4.2 - Data Storage Platform) will be used to store data objects together with their associated metadata. Additionally, all components will be monitored using the tools developed in M4.5 - Service Monitoring & Reporting. Metadata and workflow standards are an overarching concern of Task area 3 - Services that will continue to be informed by M2.2 - Data, Metadata & Workflow Standards.

Risks of implementation and mitigation approaches: Although the first funding phase established many foundational services and collaborative structures between Task area 3 - Services and the other TAs, several risks persist in the second phase. Chief among them are uneven adoption of agreed standards and the possibility that new platforms or services will not fully meet user needs. To mitigate these risks, we will closely monitor service development and compliance with our own guidelines and community standards. Active community engagement, including regular surveys and events, will help us identify emerging issues and ensure that services are continuously adapted to meet evolving requirements. Where challenges are identified, we will increase the involvement of software developers, enhance personal exchange between teams, and update requirement specifications to maintain alignment and usability.

5.4 Task area 4 - Technical Infrastructure

TA-Coordinators: UNIBI (lead), JLU (co-lead)

TA4 provides the operational foundation and technical backbone for NFDI4Microbiota, delivering the robust, sustainable infrastructure essential for all consortium activities and ensuring the reliable operation, development, and seamless integration of the computational, storage, and workflow platforms underpinning the consortium (Fig. 14). During the first funding period, we established and maintained the key infrastructure needed for NFDI4Microbiota, including coordinated operation of the federated de.NBI Cloud and large-scale virtualised compute and storage resources (M4.1 - Computational Infrastructure Operations and Software Components). Key achievements include the deployment of BiBiGrid for automated cloud-based cluster setup, SimpleVM for user-friendly cloud access, and the Aruna platform for cloud-native data storage and orchestration (M4.2 - Data Storage Platform). We also initiated an integrated platform, CloWM, for highly scalable workflow execution. CloWM couples a GUI with our flexible compute and storage infrastructure, making the consortium's workflow suite widely accessible. Together, these foundational components enable secure, scalable, and compliant delivery of the consortium's services and analytical workflows. In the next funding phase, our focus will be on consolidating and expanding these technological capabilities. We will continue to operate and govern distributed compute and storage resources for data-intensive microbiome research, including Trusted Research Environments (TREs) and cloud-based solutions for emerging technologies such as LLMs (M4.1 - Computational Infrastructure Operations and Software Components). We will advance solutions for unified, federated data storage across diverse sources (M4.2 - Data Storage Platform), and further develop CloWM as a dedicated, scalable platform for data analysis (M4.3 - Workflow Execution Platform). At the same time, we will ensure seamless interoperability with national and international infrastructures, including Base4NFDI and EOSC (M4.4 - Base4NFDI and EOSC Service Integration). Ongoing monitoring and transparent reporting (M4.5 - Service Monitoring & Reporting) will guarantee secure, efficient, and reliable operation adapted to evolving requirements. By delivering this integrated, future-proof infrastructure, TA4 ensures that all NFDI4Microbiota services remain robust, scalable, and sustainable, supporting both current operations and long-term scientific innovation across the consortium.

5.4.1 M4.1 - Computational Infrastructure Operations and Software Components

Contributors: UNIBI (lead): 27.0 PM and in-kind, JLU (co-lead): 30.0 PM, FSU: in-kind

Goals: We will provide and operate the necessary compute and storage resources for the NFDI4Microbiota services. This includes the provisioning and governance of all de.NBI Cloud Cloud resources used by NFDI4Microbiota, such as (preconfigured) virtual machines, as well as further development of cloud-related software components such as (Trusted) Research Environments ((T)RE), the BiBIGrid software for cloud-based cluster setup, and support for the local deployment of pre-trained LLMs.

Major accomplishments of first funding period: Since project launch we have operated the underlying computational infrastructure for NFDI4Microbiota via the de.NBI Cloud,

contributing approximately 630 virtual machines with more than 6000 cores, 30 TB of RAM, 350 TB of storage, an additional 500 TB of object storage, and 5 Graphics Processing Units (GPUs). We developed BiBiGrid, a tool that sets up SLURM clusters with dynamic resource provisioning in the cloud. BiBiGrid now forms a core part of the execution layer in the CloWM workflow execution platform developed in M4.3 - Workflow Execution Platform. We also began developing SimpleVM to simplify access to cloud resources; it enables fast deployment of preconfigured virtual machines and is already used in the consortium's hands-on data analysis courses and other training activities.

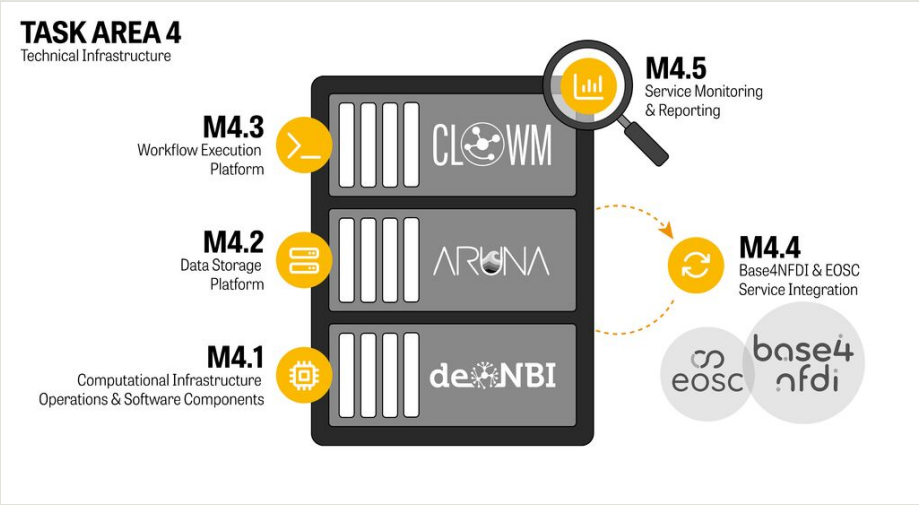

Figure 14.

Overview of Task Area 4

Planned activities for the second funding period: Operating, provisioning, and governing the compute and storage resources required by NFDI4Microbiota services is an ongoing task of this measure. As a provider of cloud-based compute and storage resources, NFDI4Microbiota will continue to draw on the federated de.NBI Cloud in line with task c) of the NFDI Expert Committee (NFDI Expert Committee 2023). To further simplify access to computational resources for demanding bioinformatics analyses, we will evolve the SimpleVM software into a stand-alone application deployable on any OpenStack cloud and compatible with any single sign-on service. The upgraded SimpleVM will function as a self-service platform offering (T)REs through pre-configured virtual machines. Beyond basic data-processing use-cases, it will support sophisticated, cluster-based computing, GPU-accelerated machine learning, and workshop scenarios. We will also offer programmatic access to dedicated, pre-trained LLMs via an API. The open-source Ollama framework will be used to deploy local LLMs, backed by special machine-learning GPUs. This will allow users to interact either via the Open WebUI for conversation or directly through the API for programmatic purposes. Alongside ongoing support for SLURM-based clusters managed by the BiBiGrid framework, we will also support self-managed Kubernetes (K8s) clusters as an alternative compute backend.

This measure also encompasses hosting and operating a Nextcloud instance for NFDI4Microbiota's central document server.

Deliverables:

- Operation of cloud services. Provisioning and governance of cloud resources
- Deployment of SimpleVM
- (T)REs
- Availability of dedicated, pre-trained LLMs for local usage.
- Provisioning of the BiBiGrid framework
- Self-managed K8s clusters available
- Operation of the consortias central Nextcloud-based document server

Table 31 lists the milestones for M4.1.

Table 31.
Milestones to be achieved in measure 4.1 - Computational Infrastructure Operations and Software Components

Milestone	Month	Description
MS4.1.1	6	Seamless access to cloud resources in place
MS4.1.2	12	Pre-trained LLMs with programmatic access in place
MS4.1.3	18	Self-managed K8s clusters available

5.4.2 M4.2 - Data Storage Platform

Contributors: JLU (lead): 120.0 PM, UNIBI (co-lead): 10.0 PM, FSU: 5.0 PM

Goals: Modern microbiology research generates vast, complex datasets that increasingly require cross-disciplinary collaboration and integrated analysis to unlock meaningful biological insights. However, today's research infrastructures remain fragmented, creating data silos that hinder knowledge discovery and reproducible science. To address these challenges, Aruna (Dieckmann et al. 2023) has been developed as a federated, cloud-native, scalable, domain-independent data-orchestration engine. It is designed to connect multiple NFDI consortia while maintaining data sovereignty and complying with each consortium's internal governance policies.

Major accomplishments of first funding period: During the first funding period, Aruna was developed as a distributed cloud-native, scalable, and domain-independent data orchestration engine in close collaboration with NFDI4Biodiversity and the FAIR Data Spaces project. It offers unified access to heterogeneous sources through an S3-compatible interface. It also offers flexible deployment capabilities, enabling individual institutions to maintain local control while participating in the ecosystem. Aruna provides a comprehensive public search index for dataset discovery and automated data processing workflows via its Hook feature, reducing manual intervention and accelerating

scientific insights. Since the release of version 2.0 in March 2024, Aruna has demonstrated remarkable success with more than 50 users from various scientific communities registering over 750 000 research objects totalling more than 15 TB. Its widespread uptake is evident in integrations with services across multiple consortia, including VirJenDB (M3.3- Databases & Data Services) from NFDI4Microbiota, Disc3D, and Visualization, Analysis and Transformation Tool (VAT) from NFDI4Biodiversity.

Planned activities for the second funding period: Throughout the first funding phase, multiple additional focus areas emerged based on user feedback, changes in the global research data landscape, and evolving community needs. One major area of focus is addressing the growing demands for scalability. Increasingly restrictive international research environments have created urgent demand for domestic data-storage solutions, pushing requirements far beyond initial projections. The increase affects not only total volume but also the sheer numbers of individual objects. Applications such as ZincDB or BakRep (M3.3 - Databases & Data Services) must now manage hundreds of millions or even billions of discrete files, datasets, and research artefacts, which require sophisticated indexing and retrieval capabilities. An equally important topic is the need for intelligent data orchestration and the preservation of data sovereignty. Fast, efficient data access is essential for effective microbial research. As datasets expand and computational loads rise, systems must intelligently position data and compute. By weighing resource needs, sensitivity, regulatory constraints and access patterns, they can decide whether to move compute to data or data to compute, accelerating scientific discovery while safeguarding data sovereignty. To facilitate a smooth transition between data storage and processing, we will intensify joint development of Aruna and CloWM (M4.3 - Workflow Execution Platform). Finally, ensuring alignment with international standards is essential for effective global collaboration. The international research community—through organisations such as the ELIXIR Network and the GA4GH—has defined standardised interfaces that enhance global service accessibility. These include data-access protocols like Data Repository Service (DRS) and Task Execution Service (TES), along with comprehensive metadata specifications such as Bioschemas and RO-Crate. Adopting these interfaces is essential for cross-border collaboration and true interoperability in the global research ecosystem. Based on identified needs and community feedback, we will transform Aruna into a fully federated, distributed peer-to-peer system that tackles the core challenges of scalability, sovereignty, and interoperability in modern microbiology research. The new architecture will enable scalability by distributing computational and storage resources across a network of fully autonomous nodes, which adheres to the task g) of the NFDI Expert Committee (NFDI Expert Committee 2023). A key innovation here is the democratisation of participation. Any institution or research group can contribute its own sovereign nodes to the federation while maintaining full control over its data and governance policies. This enables the system to expand organically as research needs grow and new institutions join the network, while also fulfilling the requirements of tasks b) and h) set out by the NFDI Expert Committee (NFDI Expert Committee 2023). To this end, we will merge Aruna's existing data and metadata components into a single, configurable application delivered through standardised, user-friendly deployment methods. We will provide comprehensive

documentation and streamlined setup processes that enable institutions to deploy and operate these federated nodes with minimal technical overhead, ensuring participation barriers remain low regardless of organisational size or technical expertise. Aruna nodes will intelligently coordinate compute-to-data and data-to-compute strategies through advanced orchestration. By weighing data locality, compute requirements, regulatory constraints and access patterns, the system will make strategic decisions about workload placement to optimise resource utilisation. This federation will turn today's fragmented research data landscape into a unified yet distributed infrastructure that preserves institutional autonomy while enabling seamless cross-disciplinary collaboration and data integration across the microbiology research community. This is in line with tasks a) and c) set out by the NFDI Expert Committee (NFDI Expert Committee 2023). To align with international standards, Aruna will provide broad support for well-established frameworks, including GA4GH protocols (DRS and TES), Bioschemas metadata specifications, and the RO-Crate packaging standard, in cooperation with M2.2 - Data, Metadata & Workflow Standards. This aligns with EOSC federation principles for cross-border data discovery while respecting local governance models. Standardised metadata formats will facilitate automated exchange, federated authentication, and AI-powered capabilities such as intelligent data discovery and automated quality assessment, significantly enhancing value extraction from federated research data. Beyond the core standards, Aruna will provide a flexible metadata extension framework allowing research communities to define domain-specific schemas—custom fields for specialised microbiological data types, experimental protocols, community-specific ontologies—without compromising interoperability.

Deliverables:

- First prototype that replaces the existing centralised architecture with a federated approach using peer-to-peer technologies
- Finalisation of necessary API and metadata format changes to support federated operation, including RO-crate support
- Integration of community interfaces for data and task distribution (DRS, TES)
- Integrated data and metadata handling components supporting transparent access, efficient transfer, replication, federated search, and cross-domain discovery
- Production-ready federated system with comprehensive usage and deployment documentation

Table 32 lists the milestones for M4.2.

5.4.3 M4.3 - Workflow Execution Platform

Contributors: UNIBI (lead): 86.0 PM and in-kind, EMBL (co-lead): 5.0 PM and in-kind, FSU: 5.0 PM, JLU: in-kind

Goals: This measure aims to provide a user-friendly, scalable, and reproducible workflow execution platform that makes best-practice data analysis readily accessible to the life

science community, even in data-intensive application scenarios. Support will be provided for a wide range of computational workflows and for seamless integration with NFDI4Microbiota infrastructure.

Table 32.
Milestones to be achieved in measure 4.2 - Data Storage Platform

Milestone	Month	Description
MS4.2.1	8	Federated prototype ready
MS4.2.2	12	API defined
MS4.2.3	24	RO-Crate and Bioschemas support integrated
MS4.2.4	30	Data & Metadata handling ready
MS4.2.5	36	TES & DRS interface ready
MS4.2.6	42	Data placement strategies available
MS4.2.7	50	Federated Aruna system ready
MS4.2.8	51	Productive operation initiated

Major accomplishments of first funding period: In the first funding phase we began developing CloWM (Göbel et al. 2024), a comprehensive and open platform that turns command-line-driven workflows into user-friendly web-based services. CloWM offers a seamless integration of (1) scientific workflows written in the Nextflow DSL (Di Tommaso et al. 2017), (2) robust data storage, (3) a highly scalable compute layer for data-intensive analysis tasks, and (4) a user-friendly interface. Since its public launch in September 2024, CloWM has garnered considerable attention from the microbiology community and other NFDI consortia (e.g., NFDI4Biodiversity). At the time of writing, the CloWM service has more than 100 registered users and hosts over 30 curated, production-ready workflows with several hundred executions. Additional workflows are under active development on the CloWM staging instance. Current workflows cover a broad spectrum of research fields and tasks, including metagenomics, human-variation analysis, infectious diseases, transcriptomics, basecalling, and phylogenomics. Some of these are exclusive to CloWM, such as the Metagenomics-Toolkit workflow (Belmann et al. 2025), a scalable and flexible workflow for complete end-to-end metagenomic data analysis, while others originate from well-known public collections like nf-core (Ewels et al. 2020) and EPI2MELABS.

Planned activities for the second funding period: In this next funding phase, we will develop the CloWM platform into an enhanced, feature-rich, mature web service accessible to all users, developers, and services within NFDI (M3.3 - Databases & Data Services, M1.4 - Connection to other NFDI Consortia), thereby fulfilling tasks g) and h) of the NFDI Expert Committee (NFDI Expert Committee 2023). To achieve this, we will integrate secure user authentication with various community Authentication and Authorization Infrastructures (AAIs) available in the NFDI by leveraging the InfraProxy

IAM4NFDI base service. We will also expand the number of curated workflows available on CloWM by (1) making further best-practice (see M2.3 - Best Practices for Data Processing) workflows developed in M3.2 - Analytical Services & Workflows available on the platform, and (2) enhancing support for existing workflows from public collections. By integrating Snakemake (Mölder et al. 2021), we will add support for a second widely used workflow management system alongside Nextflow. To embed CloWM seamlessly in the NFDI4Microbiota ecosystem, the storage platform developed in M4.2 - Data Storage Platform will be incorporated as an additional storage provider within CloWM. We also aim to support further workflow execution environments such as K8s and multiple interacting CloWM instances across different cloud sites (data centers), optimising hardware use, especially costly GPUs, while reducing data transfers. Because many workflows run for hours or days on large datasets, we will add real-time execution monitoring, giving users and developers detailed status views and better debugging. To enhance the reliability of workflow execution, we will implement a generic parameter interface with strict pre-launch validation of workflow parameters. To enable automated post-processing of computed workflow results by downstream data analysis services, we will implement WebHooks for event-based messaging. Finally, to improve discoverability of workflows on CloWM, we will utilise (Bioschemas) metadata markup, enhancing search-engine indexing and enabling automatic registration in public registries such as the ELIXIR WorkflowHub (Gustafsson et al. 2024) (M1.5 - Connection to International Partners). To grow the CloWM user and workflow developer communities, we will establish CIGs for these communities in close collaboration with M1.3 - Community Engagement and create training materials specifically designed for different target audiences. This will meet the requirements of task b) of the NFDI Expert Committee (NFDI Expert Committee 2023). Our training programme will offer regular workshops, delivering introductory sessions for general platform users and advanced courses for technically experienced workflow developers.

Deliverables:

- Continuous improvement of CloWM and further development of the platform into a mature web service accessible to everyone within NFDI
- Availability of new best-practice analytical workflows developed on the CloWM platform
- Seamless integration of the NFDI4Microbiota data orchestration platform Aruna
- Support for workflows written in Snakemake DSL
- Real time workflow monitoring for detailed status assessment and debugging
- Generic workflow parameter interface with enhanced parameter validation
- Event-based messaging for automated workflow post-execution tasks.
- Establishment of CloWM-related CIGs and target-audience-specific training.

Table 33 lists the milestones for M4.3.

Table 33.
Milestones to be achieved in measure 4.3 - Workflow Execution Platform

Milestone	Month	Description
MS4.3.1	3	Support for NFDI community AAIs and event-based messaging in place
MS4.3.2	12	Simplified workflow registration process for developers in place
MS4.3.3	18	Establishment of CIGs. Training materials available for different target audiences (CloWM users and CloWM workflow developers)
MS4.3.4	24	Seamless integration of the Aruna data storage platform from M4.2 - Data Storage Platform into CloWM implemented
MS4.3.5	36	Workflow real-time monitoring and extended parameter validation established
MS4.3.6	48	Full integration of the Snakemake workflow language in place
MS4.3.7	60	Support for multiple CloWM instances with synchronisation of persistent resources in place. Support for K8s as an alternative execution engine in place

5.4.4 M4.4 - Base4NFDI and EOSC Service Integration

Contributors: JLU (lead): 30.0 PM, UNIBI (co-lead): 2.0 PM, EMBL: 5.0 PM, FSU: in-kind

Goals: This measure aims to support the integration of NFDI4Microbiota services with Base4NFDI by providing guidance, technical assistance, and coordination for adopting common infrastructure components. In parallel, it will evaluate and align the services with EOSC requirements, ensuring interoperability and visibility within EOSC. Close collaboration with relevant NFDI and EOSC partners will keep the work in step with evolving technical and policy standards. Together, these efforts will enhance data accessibility, lower technical barriers for researchers, and strengthen the overall research infrastructure ecosystem by enabling seamless data sharing and collaboration across platforms.

Major accomplishments of first funding period: Because this measure is new, there are no achievements from the first funding period.

Planned activities for the second funding period: This measure combines two activities that reinforce the technical and organisational foundations of federated research services. First, it provides comprehensive support for the integration of NFDI Base4NFDI services within the consortium (M3.3 - Databases & Data Services). The measure team will serve as a primary contact point for modules such as IAM4NFDI, PID4NFDI, TS4NFDI, and DMP4NFDI (as ZB MED is project lead of KGI4NFDI and RDMTraining4NFDI they are already deeply integrated). We will offer guidance, documentation, and hands-on assistance to consortium partners and serve as an interface for coordination with Base4NFDI developers and external consortia, working in close collaboration with the Base4NFDI service stewards, which is in line with tasks b), e)

and g) of the NFDI Expert Committee (NFDI Expert Committee 2023). Second, the measure will provide strategic guidance and coordinated support to ensure that NFDI4Microbiota services achieve full compatibility with established international standards and frameworks. This includes offering technical consultation and best-practice recommendations for implementing OpenAPI specifications, achieving FAIR Digital Object (FDO) compliance, and meeting EOSC interoperability requirements. The work of NFDI4Microbiota will facilitate international cooperation, via existing and long-standing research and infrastructure collaborations in field microbiology. We will contribute to the establishment of the National Node within NFDI by providing our services. As a knowledge hub for consortium partners, we will provide documentation, training materials, and ongoing advisory support to facilitate alignment with ELIXIR infrastructure standards and European research ecosystem requirements. Additionally, the measure will provide implementation support for privacy-compliant data-analysis environments such as TREs, working in close collaboration with the simplified infrastructure provisioning activities of M4.1 - Computational Infrastructure Operations and Software Components, which is in line with task h) of the NFDI Expert Committee (NFDI Expert Committee 2023).

Deliverables:

- Integration support for IAM4NFDI, PID4NFDI, TS4NFDI, and DMP4NFDI
- Annual workshop(s)
- Creation of internal guidelines

Table 34 lists the milestones for M4.4.

Table 34.
Milestones to be achieved in measure 4.4 - Base4NFDI and EOSC Service Integration

Milestone	Month	Description
MS4.4.1	6	Support structure and contact point for Base4NFDI established
MS4.4.2	12	Whitepaper for internal guidelines available
MS4.4.3	13	Support phase starting

5.4.5 M4.5 - Service Monitoring & Reporting

Contributors: EMBL (lead): 10.0 PM, ZB MED (co-lead): 10.0 PM, FSU: in-kind, JLU: in-kind, ZB MED: in-kind

Goals: The goal of this measure is to evaluate and quantify the impact of NFDI4Microbiota within the community by monitoring a set of established KPIs.

Major accomplishments of first funding period: We established different mechanisms to systematically track KPIs relevant to service usage (M3.3 - Databases & Data Services, M3.1 - Central Web Portal, M3.2 - Analytical Services & Workflows) among others with a

dedicated Matomo instance. These include, but are not limited to, user numbers, frequency of access, dataset uploads/downloads, and other critical usage metrics. The collected data provide actionable insights into how the services (from M3.2 - Analytical Services & Workflows, M3.3 - Databases & Data Services, M4.3 - Workflow Execution Platform, and others linked in the M3.1 - Central Web Portal) are accepted and utilised by the target communities. These metrics follow DFG guidelines, allowing us to highlight successes and identify areas needing improvement.

Planned activities for the second funding period: While support services for KPI monitoring are in place from the first funding phase, they require ongoing maintenance to ensure accurate and up-to-date reporting. Regular dashboards and narrative reports will foster transparency and enable data-driven service optimisation. This work package will also establish standardised reporting protocols to the NFDI and DFG to ensure that stakeholders and funders are kept informed about service performance and project outcomes throughout the project lifecycle.

Deliverables:

- Operation of Matomo service
- KPI collection for workflows running on the CloWM platform

Table 35 lists the milestones for M4.5.

Table 35.
Milestones to be achieved in measure 4.5 - Service Monitoring & Reporting

Milestone	Month	Description
MS4.5.1	6	Reporting protocols formalised
MS4.5.2	9	KPI dashboard operational
MS4.5.3	12	Annual performance report published
MS4.5.4	24	Annual performance report published
MS4.5.5	48	Annual performance report published
MS4.5.6	60	Annual performance report published

5.4.6 Overall schedule, contribution, risks and dependencies

Fig. 15 illustrates the overall schedule for TA4.

Contributions: TA4 remains the technical backbone of NFDI4Microbiota, providing the infrastructure that underpins all consortium activities. In this highly specialised area, five out of ten co-applicants (UNIBI, JLU, EMBL, FSU Jena, and ZB MED) contribute directly, reflecting the technical focus on cloud operations, federated storage, workflow services, and monitoring. UNIBI and JLU jointly coordinate cloud operations and storage

infrastructure, supported by EMBL and FSU Jena, with JLU leading the federated data storage platform and UNIBI driving workflow and cloud services. ZB MED together with EMBL, coordinates service monitoring and reporting. All partners will continue to work closely with other task areas and user communities to ensure the infrastructure meets evolving scientific needs.

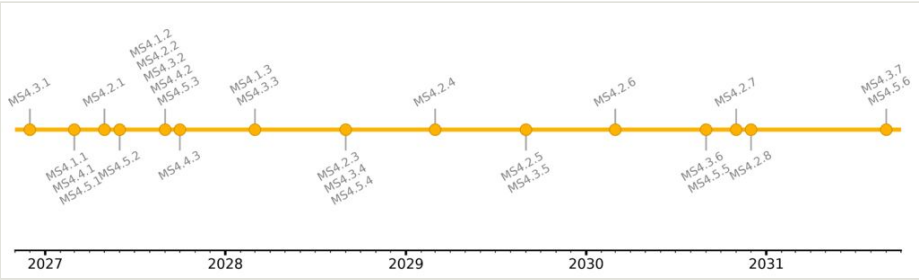

Figure 15.

Overall schedule Task Area 4

Dependencies: TA4 is essential for deploying and operating all technical services and tools within the consortium. Ongoing integration with national and international infrastructures, along with regular feedback from other task areas, is critical to maintain interoperability and support platform evolution.

Risks of implementation and mitigation approaches: Key risks include possible outages in cloud or storage services, which will be mitigated by building in greater redundancy and supporting federated deployments. Integration challenges are anticipated, but will be managed by dedicating skilled software developers and engineers to support both users and tool development. Achieving alignment with other NFDI consortia may be complex, so active joint planning and a clear definition of technical priorities will help keep efforts coordinated.

5.5 Task area 5 - Coordination & Communication

TA-Coordinators: ZB MED (lead), EMBL (co-Lead)

TA5 underpins and sustains the organisational framework for NFDI4Microbiota, ensuring efficient coordination, transparent governance, and sustainable management across the entire consortium (Fig. 16). During the first funding phase, we implemented a robust governance structure, including functional decision-making and advisory bodies, clear project governance processes (M5.1 - Project Governance), and effective financial controlling and reporting (M5.2 - Project Financial Controlling & Reporting). We streamlined interaction and collaboration between partners and governing bodies through a GDPR-compliant internal platform for document sharing (Nextcloud), scheduling, and meetings. Regular BoD meetings and close milestone tracking ensured timely progress, maintained flexibility, and enabled rapid adaptation to emerging needs. Looking ahead to the second funding phase, we will further strengthen and refine these

core structures. The AU will continue preparing activity and financial reports, allocating resources transparently, and documenting all processes. Strategic and legal issues will remain closely aligned with the NFDI Directorate, and any required changes to work or financial plans will be coordinated transparently via the GA. Flexible adaptation and growth (M5.3 - Dynamic Adaptation & Growth) will be supported by dynamic mechanisms that respond to evolving community needs, best practices, and scientific advances. By maintaining and further developing these key governance and coordination mechanisms, TA5 will provide the stable, transparent, and responsive framework required for all NFDI4Microbiota activities.

Figure 16.
Overview of Task Area 5

5.5.1 M5.1 - Project Governance

Contributors: ZB MED (lead): 85.5 PM, HZI (co-lead): in-kind, UMR: in-kind, ZB MED: in-kind

Goals: To ensure robust governance across the consortium and support its internal bodies.

Major accomplishments of first funding period: We established all the consortium's internal bodies (see 3.4 Organisational structure and viability) and adopted a CA, providing a formal legal framework that has effectively supported every TA.

Planned activities for the second funding period: We will maintain these governance activities and aim to make more of our strategic documents and process descriptions available for reuse. This might also include significant revisions to the plan due to the proposed reorganisation of the NFDI structure (German Science And Humanities Council

2025, Rat für Informationsinfrastrukturen 2025) also in close collaboration with M1.4 - Connection to other NFDI Consortia.

Deliverables:

- KPI-based decision framework (based on M4.5 - Service Monitoring & Reporting)
- NFDI governance structure plan (with all consortia) and transition plan based on this for NFDI4Microbiota
- Updated CA
- Public strategic documents and process description collection
- Organisation of GA meetings
- Event organisation (e.g. annual conference together with M1.3 - Community Engagement)

Table 36 lists the milestones for M5.1.

Table 36.
Milestones to be achieved in measure 5.1 - Project Governance

Milestone	Month	Description
MS5.1.1	6	First SAB meeting of the second funding phase
MS5.1.2	9	Sustainability and transition plan
MS5.1.3	12	Annual report 2026
MS5.1.4	24	Annual report 2027
MS5.1.5	36	Annual report 2028
MS5.1.6	48	Annual report 2029
MS5.1.7	60	Final report

5.5.2 M5.2 - Project Financial Controlling & Reporting

Contributors: ZB MED (lead): 28.5 PM, EMBL (co-lead): 15.0 PM, HZI: in-kind, UMR: in-kind

Goals: To ensure robust financial coordination and controlling across the consortium, supporting transparent reporting to both internal stakeholders and funders.

Major accomplishments of first funding period: In the first funding phase, we established efficient financial guidance, justified resource allocations, and produced regular summary financial reports. The legal basis for financial controlling and reporting was formalised through the CA. Annual project costs were forecasted, applications to carry forward unspent funds were submitted as needed, and quarterly financial requirements were collected from all co-applicants. Funds were requested from the DFG and distributed promptly to all partners.

Planned activities for the second funding period: The financial controlling team will maintain these core responsibilities, working closely with the BoD and AU to safeguard transparency, compliance, and efficient resource management.

Deliverables:

- Financial management plan
- Finances forecasting
- Budget allocation and justification report
- Sustainability cost model

Table 37 lists the milestones for M5.2.

Table 37.
Milestones to be achieved in measure 5.2 - Project Financial Controlling & Reporting

Milestone	Month	Description
MS5.2.1	4	Financial report 2026
MS5.2.2	16	Financial report 2027
MS5.2.3	28	Financial report 2028
MS5.2.4	40	Financial report 2029
MS5.2.5	52	Financial report 2030
MS5.2.6	60	Financial report 2031

5.5.3 M5.3 - Dynamic Adaptation & Growth

Contributors: ZB MED (lead): 14.25 PM, EMBL (co-lead): 10.0 PM, Flex Funds: 91.0 PM, HZI: in-kind

Goals: To ensure that NFDI4Microbiota continuously refines its strategy and services to meet community needs, adopts innovative methodologies, and adheres to best practices, all while sustaining the consortium's infrastructure.

Major accomplishments of first funding period: We established a flexible funding mechanism (Flex Fund) that enabled the consortium to respond dynamically to new scientific challenges and community requirements. Annual competitive project calls, evaluated through external peer review and coordinated by a PRB nominated by the SAB, allowed us to select and support projects that best advanced the mission of NFDI4Microbiota. This flexible scheme also covered major partner contributions and unforeseen technical obstacles.

Planned activities for the second funding period: We will continue to use Flex Funds to expand and adapt the consortium's services in a dynamic, demand-driven way,

maintaining the existing call and review process to remain responsive to the evolving needs of the community.

Deliverables:

- Flex Fund call organisation
- Candidate list for new PRB

Table 38 lists the milestones for M5.3.

Table 38.
Milestones to be achieved in measure 5.3 - Dynamic Adaptation & Growth

Milestone	Month	Description
MS5.3.1	10	Annual FlexFund call 2028
MS5.3.2	22	Annual FlexFund call 2029
MS5.3.3	31	Annual FlexFund report meeting for the 2028 round
MS5.3.4	34	Annual FlexFund call 2030
MS5.3.5	43	Annual FlexFund report meeting for the 2029 round
MS5.3.6	46	Annual FlexFund call 2031 (short runtime)
MS5.3.7	55	Annual FlexFund report meeting for the 2031 round

5.5.4 Overall schedule, contribution, risks and dependencies

Fig. 17 illustrates the overall schedule for TA5.

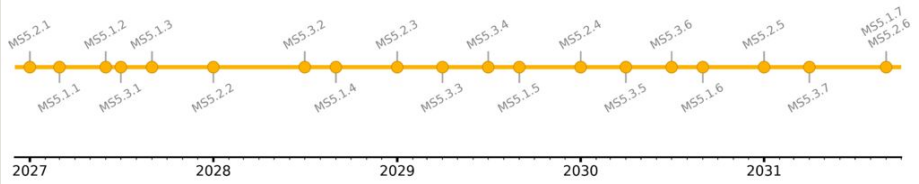

Figure 17.
Overall schedule Task Area 5

Contributions: Most governance and administrative structures for TA5 were established during the first funding phase and now provide a stable foundation for the consortium. In the second phase, ZB MED will lead EMBL, HZI, UMR, and RWTH to maintain and develop these core functions with support from the AU. This collaborative approach ensures effective project management, financial coordination, transparent reporting, and flexible support for new initiatives as the consortium's needs evolve.

Dependencies: TA5 relies on timely cooperation and information exchange with all other TAs for progress monitoring and financial oversight, and requires close alignment with the DFG and NFDI Directorate in regard to reporting and compliance with legal and strategic requirements.

Risks of implementation and mitigation approaches: Risks include the need to adapt activities or budgets, which may occasionally lead to conflicting interests. These situations are managed through transparent communication among the partners involved, with escalation to the GA if needed. Should deliverables be delayed or quality fall short, tasks and resources can be reassigned to maintain continuity.

6 Additional Aspects

6.1 Equity and diversity

NFDI4Microbiota follows the NFDI Code of Conduct, which emphasises that diversity is a strength and essential for success. The code forbids any form of sexism, racism, insults, bullying, or sexual harassment, and promotes cooperation based on mutual respect. Between July 2024 and June 2025, an average of 40% of people working for NFDI4Microbiota were women, and 42% were individuals with a migrant background, representing 13 nationalities other than German. All (co-)applicant institutions have implemented measures to ensure equal opportunities and workforce diversity in their recruitment processes. These institutions hold various certifications demonstrating their commitment to these principles, such as ”Total Equality,” ”Shaping Diversity,” ”Charta der Vielfalt,” ”Weltoffenes Thüringen,” and ”family-friendly university,” among others. Additionally, they collaborate with coalitions advocating for diversity in science, including Equality, Diversity, and Inclusion in Science and Health (EDIS). Beyond these fundamental principles for promoting diversity within the consortium, further measures have been implemented. In coordination with the Ethical, Legal, and Social Aspects (ELSA) cross-section and the ”Inclusion in RDM” subgroup, NFDI4Microbiota has taken concrete steps to enhance diversity.

6.2 Further comments

The authors acknowledge the use of LLMs to phrase texts based on original notes, ideas, and concepts. Full responsibility for the content remains with the authors.

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Belmann P, Osterholz B, Kleinboelting N, Puehler A, Schlueter A, Sczyrba A (2025) Metagenomics-Toolkit: the flexible and efficient cloud-based metagenomics workflow featuring machine learning-enabled resource allocation. NAR Genomics and Bioinformatics 7 (3). https://doi.org/10.1093/nargab/lqaf093
- Bernard L, Brück M, Busse C, Engel J, Eufinger J, Ewert F, Fluck J, Förstner K, Fürst J, Gauza H, Getzlaff K, Glöckner FO, Hunold J, Koepler O, Krooß K, Lindstädt B, McHardy A, Mehrtens H, Rey-Mazon E, Schmidt M, Schober I, Schröter A, Stegle O, Steinbeck C, Steinhart F, von Suchodoletz D, Weidtkamp-Peters S, Wendt J, Wetzker C (2025) Memorandum of Understanding of NFDI consortia from Earth-, Chemical and Life Sciences to support a network called the Geo-Chem-Life Science Helpdesk Cluster. Zenodo. doi: 10.5281/ZENODO.15065069. URL: https://zenodo.org/doi/10.5281/zenodo.15065069
- Clavel T, Faber F, Groussin M, Haller D, Overmann J, Pauvert C, Poyet M, Selkrig J, Stecher B, Typas A, Vehreschild MGT, Westermann A, Wylensek D, Maier L (2025) Enabling next-generation anaerobic cultivation through biotechnology to advance functional microbiome research. Nature Biotechnology 43 (6): 878-888. https://doi.org/10.1038/s41587-025-02660-6
- Dieckmann MA, Beyvers S, Hochmuth J, Rehm A, Förster F, Goesmann A (2023) The Aruna Object Storage: A Distributed Multi Cloud Object Storage System for Scientific Data Management. Proceedings of the Conference on Research Data Infrastructure 1 https://doi.org/10.52825/cordi.v1i.404
- Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C (2017) Nextflow enables reproducible computational workflows. Nat. Biotechnol. 35 (4): 316-319. https://doi.org/10.1038/nbt.3820
- Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, Garcia MU, Di Tommaso P, Nahnsen S (2020) The nf-core framework for community-curated bioinformatics pipelines. Nat. Biotechnol. 38 (3): 276-278. https://doi.org/10.1038/s41587-020-0439-x
- FAIR Data Maturity Model Working Group (2020) FAIR Data Maturity Model. Specification and Guidelines. Zenodo. doi: 10.15497/rda00050. https://doi.org/10.15497/rda00050
- Förstner K, Becker A, Blom J, Bork P, Clavel T, Dieckmann M, Goesmann A, Götz B, Gübitz T, Hufsky F, Jünemann S, Körner M, Marz M, Da Rocha UN, Overmann J, Pühler A, Rebholz-Schuhmann D, Sczyrba A, Stoye J, Vandendorpe J, Van Rossum T, McHardy A (2023) NFDI4Microbiota – national research data infrastructure for microbiota research. Research Ideas and Outcomes 9 https://doi.org/10.3897/rio.9.e110501
- Förstner K (2024) NFDI4Microbiota - Making Microbiology Data FAIR and Open. Zenodo. doi: 10.5281/zenodo.13946110. https://doi.org/10.5281/zenodo.13946110
- Förstner K, Ebert B, Glöckner F, von Suchodoletz D, McHardy AC, Mühlhaus T (2024) Memorandum of Understanding between NFDI4Biodiversity, NFDI4Microbiota and DataPlant. Zenodo https://doi.org/10.5281/zenodo.13880944
- German Science And Humanities Council (2025) Strukturevaluation der Nationalen Forschungsdateninfrastruktur (NFDI). Wissenschaftsrat https://doi.org/10.57674/WCDC-6D36

- Göbel D, Stoye J, Sczyrba A, Beckstette M (2024) The Cloud-based Workflow Manager (CloWM) - An integrated platform for highly scalable workflow execution. Zenodo. doi: 10.5281/zenodo.14039069. https://doi.org/10.5281/zenodo.14039069
- Gustafsson OJR, Wilkinson SR, Bacall F, Pireddu L, Soiland-Reyes S, Leo S, Owen S, Juty N, Fernández JM, Grüning B, Brown T, Ménager H, Capella-Gutierrez S, Coppens F, Goble C (2024) WorkflowHub: a registry for computational workflows. Scientific Data https://doi.org/10.1038/s41597-025-04786-3
- Huttenhower C, Finn R, McHardy AC (2023) Challenges and opportunities in sharing microbiome data and analyses. Nature Microbiology 8 (11): 1960-1970. https://doi.org/10.1038/s41564-023-01484-x
- Keith DA, Ferrer-Paris JR, Nicholson E, Bishop MJ, Polidoro BA, Ramirez-Llodra E, Tozer MG, Nel JL, Mac Nally R, Gregr EJ, others (2022) A function-based typology for Earth's ecosystems. Nature 610 (7932): 513-518. https://doi.org/10.1038/s41586-022-05318-4
- Klemm P (2024) Provenance Guidelines for Workflow and Tool Developers. Zenodo. doi: 10.5281/zenodo.13627510. https://doi.org/10.5281/zenodo.13627510
- Koblitz J, Halama P, Spring S, Thiel V, Baschien C, Hahnke RL, Pester M, Overmann J, Reimer LC (2023) MediaDive: the expert-curated cultivation media database. Nucleic acids research 51 https://doi.org/10.1093/nar/gkac803
- Koblitz J, Kawashima S, Hidaka K, Okabeppu Y, Onoda S, Otsuka R, Ichikawa N (2024) Expanding data on cultivation media and microbial traits. Center for Open Science. doi: 10.37044/osf.io/qxpn8. https://doi.org/10.37044/osf.io/qxpn8
- Koblitz J, Reimer LC, Pukall R, Overmann J (2025) Predicting bacterial phenotypic traits through improved machine learning using high-quality, curated datasets. Communications biology 8 https://doi.org/10.1038/s42003-025-08313-3
- Lissin A, Schober I, Witte J, Lüken H, Podstawka A, Koblitz J, Bunk B, Dawyndt P, Vandamme P, de Vos P, Overmann J, Reimer LC (2025) StrainInfo - The Central Database for Linked Microbial Strain Identifiers. Database https://doi.org/10.1093/database/baaf059
- Meyer F, Fritz A, Deng Z, Koslicki D, Lesker TR, Gurevich A, Robertson G, Alser M, Antipov D, Beghini F, Bertrand D, Brito J, Brown CT, Buchmann J, Buluç A, Chen B, Chikhi R, Clausen PLC, Cristian A, Dabrowski PW, Darling A, Egan R, Eskin E, Georganas E, Goltsman E, Gray M, Hansen LH, Hofmeyr S, Huang P, Irber L, Jia H, Jørgensen TS, Kieser S, Klemetsen T, Kola A, Kolmogorov M, Korobeynikov A, Kwan J, LaPierre N, Lemaitre C, Li C, Limasset A, Malcher-Miranda F, Mangul S, Marcelino V, Marchet C, Marijon P, Meleshko D, Mende D, Milanese A, Nagarajan N, Nissen J, Nurk S, Oliker L, Paoli L, Peterlongo P, Piro V, Porter J, Rasmussen S, Rees E, Reinert K, Renard B, Robertsen EM, Rosen G, Ruscheweyh H, Sarwal V, Segata N, Seiler E, Shi L, Sun F, Sunagawa S, Sørensen SJ, Thomas A, Tong C, Trajkovski M, Tremblay J, Uritskiy G, Vicedomini R, Wang Z, Wang Z, Wang Z, Warren A, Willassen NP, Yelick K, You R, Zeller G, Zhao Z, Zhu S, Zhu J, Garrido-Oter R, Gastmeier P, Hacquard S, Häußler S, Khaledi A, Maechler F, Mesny F, Radutoiu S, Schulze-Lefert P, Smit N, Strowig T, Bremges A, Sczyrba A, McHardy AC (2022) Critical Assessment of Metagenome Interpretation: the second round of challenges. Nature Methods 19 (4): 429-440. https://doi.org/10.1038/s41592-022-01431-4
- Meyer F, Robertson G, Deng Z, Koslicki D, Gurevich A, McHardy AC (2025) CAMI Benchmarking Portal: online evaluation and ranking of metagenomic software. Nucleic Acids Research https://doi.org/10.1093/nar/gkaf369

- Minamiyama Y, Ikeuchi U, Tanabe K, Yuki K, Hayashi K, Aoki T (2025) Maturity Model for Organizational Research Data Management Services. Data Science Journal https://doi.org/10.5334/dsj-2025-018
- Mölder F, Jablonski KP, Letcher B, Hall MB, Tomkins-Tinch CH, Sochat V, Forster J, Lee S, Twardziok SO, Kanitz A, Wilm A, Holtgrewe M, Rahmann S, Nahnsen S, Köster J (2021) Sustainable data analysis with Snakemake. F1000Res. 10 https://doi.org/10.12688/f1000research.29032.2
- NFDI (2022) NFDI InfraTalk: Current developments in the global Genome Commons and challenges (16 May 2022) [Video]. YouTube. URL: https://www.youtube.com/watch?v=odNBeqav2eQ&t=1806s
- NFDI4Biodiversity (2026) Biodata Interest Group Memorandum of Understanding. https://www.nfdi4biodiversity.org/en/mou-biodata-group/
- NFDI Expert Committee (2023) Key points for the second funding period of the NFDI consortia. DFG URL: https://www.dfg.de/resource/blob/204398/eckpunkte-zweite-foerderphase-en.pdf
- Pauvert C (2023) maldipickr: Dereplicate and Cherry-Pick Mass Spectrometry Spectra. doi: 10.32614/CRAN.package.maldipickr. URL: https://CRAN.R-project.org/package=maldipickr
- Pauvert C, Magel M (2024) Workshop Biometadata-02: How to describe biological data? A primer to a FAIR approach for now and the future. doi: 10.5281/zenodo.11527597. URL: https://zenodo.org/records/11527597
- Pauvert C (2025) ClavelLab/genome-assembly: trigger release zenodo. Zenodo. doi: 10.5281/zenodo.15172576. URL: https://zenodo.org/records/15172576
- Podlesny D (2025) Borkflows: Bork Group Workflows on CloWM. Zenodo. doi: 10.5281/zenodo.14801169. https://doi.org/10.5281/zenodo.14801169
- Poveda L, Farrell G, Tosatto S, Zahn M, Ruch P, Gobeill J, Waterhouse R, Dessimoz C (2025) The missing link in FAIR data policy: data resources. Zenodo. doi: 10.5281/zenodo.15724104. https://doi.org/10.5281/zenodo.15724104
- Rat für Informationsinfrastrukturen (2025) Leistung in Verantwortung. Zur Zukunft der wissenschaftlichen Informationsinfrastrukturen in Deutschland. URL: https://rfii.de/?p=12040
- Ritsch M, Cassman NA, Saghaei S, Marz M (2023) Navigating the Landscape: A Comprehensive Review of Current Virus Databases. Viruses 15 (9). https://doi.org/10.3390/v15091834
- Schober I, Koblitz J, Sardà Carbasse J, Ebeling C, Schmidt ML, Podstawka A, Gupta R, Ilangovan V, Chamanara J, Overmann J, Reimer LC (2025) BacDive in 2025: the core database for prokaryotic strain data. Nucleic Acids Research 53 https://doi.org/10.1093/nar/gkae959
- Thölken C, Vandendorpe J, Magel M (2024) NFDI4Microbiota Data Management Plan (DMP) Template. Zenodo. doi: 10.5281/zenodo.13628589. https://doi.org/10.5281/zenodo.13628589
- Tubbesing T, Schlüter A, Sczyrba A (2025) subMG automates data submission for metagenomics studies. BioData Mining 18 (1). https://doi.org/10.1186/s13040-025-00453-w
- Vailati-Riboni M, Palombo V, Loor J (2017) What Are Omics Sciences? In: Ametaj B (Ed.) Periparturient Diseases of Dairy Cows: A Systems Biology Approach. [ISBN 978-3-319-43033-1]. https://doi.org/10.1007/978-3-319-43033-1_1

- Van Brabant B, Dawyndt P, De Baets B, De Vos P (2006) A Knuckles-and-Nodes Approach to the Integration of Microbiological Resource Data. On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops. [ISBN 978-3-540-48272-7].
- Vandendorpe J, Lindstädt B (2023) Bringing discipline specificity to a generic research data management workshop concept. German Medical Science (GMS) https://doi.org/10.3205/MBI000557
- Vandendorpe J, Lindstädt B, Hellmann L (2024a) Workshop on Electronic Lab Notebooks (ELNs). Zenodo. doi: 10.5281/zenodo.13318741. https://doi.org/10.5281/zenodo.13318741
- Vandendorpe J, Adam B, Wilbrandt J, Lindstädt B, Förstner KU (2024b) Ten simple rules for implementing electronic lab notebooks (ELNs). PLoS Comput. Biol. 20 (6). https://doi.org/10.1371/journal.pcbi.1012170
- Vandendorpe J, Schober I, Magel M (2025) NFDI4Microbiota Helpdesk. Zenodo. doi: 10.5281/zenodo.15181009. https://doi.org/10.5281/zenodo.15181009
- Whyte A, Green D, Avanço K, Di Giorgio S, Gingold A, Horton L, Koteska B, Kyprianou K, Prnjat O, Rauste P, Schirru L, Sowinski C, Torres Ramos G, van Leersum N, Lazzeri E (2024) Minimum Viable Skillsets - Catalogue of Open Science Career Profiles. https://doi.org/10.5281/zenodo.11469300
- Wilbrandt J, Vandendorpe J, Grünwald K, Magel M (2025) NFDI4Microbiota World Café on Electronic Lab Notebooks (ELNs). Zenodo. doi: 10.5281/zenodo.15274473. https://doi.org/10.5281/zenodo.15274473
- Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJG, Groth P, Goble C, Grethe JS, Heringa J, 't Hoen PAC, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone S, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B (2016) The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3 (1). https://doi.org/10.1038/sdata.2016.18
- ZB MED - Informationszentrum Lebenswissenschaften, Vandendorpe J, Hellmann L (2024) Demo with eLabFTW. Published: NFDI4MicrobiotaDOI: 10.5446/68306. https://doi.org/10.5446/68306