National Research Data Infrastructure for Personal Health Data (NFDI4Health)

Spokesperson: Prof. Dr. Juliane Fluck, Head of Department Knowledge Management, ZB MED Information Centre Life Science, fluck@zbmed.de

Research area and data: Epidemiological, clinical and public health studies are generally highly standardised and well documented with structured, quality-checked and curated data. Both, public health and prospective clinical trial data have the advantage that they are collected according to well-documented protocols with deep and purpose-specific phenotyping and structured data types. These types of studies allow to investigate, e.g., clinical outcomes, efficacy of interventions, adverse effect profiles, prognostic factors and biomarker profiles. Especially, during the resent pandemic situation, we face the need for sharing, standardising and harmonising data for fast decision-making.

However, across these studies, (1) data are usually not harmonised or interoperable, (2) metadata descriptions are not easily accessible, (3) findability of the resources is hampered, (4) opportunities for access are currently very restricted which (5) makes it difficult to apply Artificial Intelligence (AI) / big data approaches. The overarching goal of NFDI4Health is to support its clinical and epidemiological research community in the best possible way to share their data with the user community in agreement with data protection/privacy regulations and ethics principles and, in the interest of improving population health, create new data analytics opportunities within the German NFDI.

Data and Community: NFDI4Health is targeting individual health data that are generated in clinical trials, epidemiological cohorts and public health surveillance.

NFDI4Health aims to create the most comprehensive inventory of German epidemiological, public health and clinical trial data. NFDI4Health will build a centralised data catalogue with elaborate search functionalities, sophisticated data access management, and a data analysis toolbox, while respecting stringent requirements for privacy concerning personal health data. Standardisation services will ensure a high degree of interoperability. Use cases covering prototypical study types and areas of research will show the feasibility of a harmonised implementation of all infrastructures, tools and services in accordance with our user communities.

As a major goal, NFDI4Health, together with the epidemiological and clinical trial communities, aims at developing a joint Metadata Repository (MDR) and complementing this metadata resource, whenever possible, with terminologies emerging in the context of national and international initiatives. The current Covid-19 pandemic highlights the necessity of such common resources and also the possibility

to reach the above advocated aims for defined use cases. NFDI4Health partners are already participating in these efforts in national and international co-operations.

To open up new opportunities for data analyses in the interest of improving health, NFDI4Health main **objectives** are:

(1) to enable findability of and access to structured health data from clinical trials, epidemiological studies, disease registries, administrative health databases and public health surveillance in Germany;
(2) to implement a health data framework for centralised searching and accessing existing decentralised epidemiological/clinical trial data infrastructures;
(3) to facilitate data sharing, record linkage, harmonised data quality assessments, federated analyses of personal health data;
(4) to enable the development and deployment of new, machine-processable consent mechanisms and innovative data access services;
(5) to support cooperation between clinical trial research, epidemiological and public health communities;
(6) to foster interoperability of currently fragmented IT solutions related to metadata repositories, cohort browsing, data quality and harmonisation;
(7) to develop business models to secure sustainability of structures and services.

NFDI4Health will provide the following main services:

1. **A *Central Search Hub* for making data discoverable and accessible,** where data holders can publish their (meta) data and users can search in data catalogues, data properties and both get data quality overviews.
2. **A *Central Access Point (CAP)* and distributed *Local Access Points (LAP)* to simplify data use request and access,** where catalogues of data items and quality properties allow for central data selection. The CAP supports the access to the LAPs and communicates with heterogeneous access governance structures.
3. **Distributed data analysis environments** will be set up based on the frameworks ***DataSHIELD*** and ***Personal Health Train***. For defined use cases, the respective local data access points will be connected within these frameworks and methods for distributed data analysis will be further developed including AI applications.

Due to the highly sensitive person-related data, we face specific challenges for our research area because of *data privacy* and *ethical principles.* That is, permission has to be checked accounting for privacy requirements before data can be shared. Thus, *data access* is highly regulated and restricted, which means that a *distributed research infrastructure* has to be established which allows for *distributed data analysis*. We will also need to change the *culture of data sharing* even for metadata, where sharing would be possible but is still not common. In addition, we need to promote closer cooperation of epidemiological, public health and medical communities.

Data management experience: NFDI4Health brings together longstanding experience in research data management (e.g. publication infrastructure for research data, data search) with the specific knowledge of standards and capabilities required for making research data FAIR (e.g. metadata standards, ontologies, data provenance). NFDI4Health data holders have established high IT standards, sound use and access procedures and high standards of data protection. NFDI4Health partners are integrated in a broad network of (inter)national experts/initiatives especially in the field of standardisation (e.g. ISO/TC 215 Health Informatics, ISO/TC 276 Biotechnology, CDISC, HL7, GA4GH, OHDSI, IHE, LOINC).

NFDI4Health intends to work with various partners and infrastructures mostly through its co-applicant networks. Many of them are already associated with NFDI4Health as participants (currently 46 participants). Furthermore, national and international initiatives and societies in the medical domain support NFDI4Health (37 letters of support).

Interfaces to other NFDI consortia: NFDI4Health seeks cooperation with all other life science NFDI consortia and with consortia dealing with person-related data beyond the health domain. With GHGA, we have agreed on a close partnership in addressing processes of patient identification, record linkage, standardisation and common services. NFDI4Health will also coordinate its efforts together with KonsortSWD in addressing challenges that are related to making sensitive cohort and survey data reusable and interoperable. Further cooperation will be established with NFDI4Med, NFDI-Neuro, NFDI4NeuroFunction, NFDI4Chem, NFDI4Microbiota, NFDI4Agri, NFDI4Earth, NFDI4BioDiversity and NFDI4NanoSafety.

Cross-cutting topics: NFDI4Health will contribute to all cross-cutting topics that are addressed in the **Berlin Declaration** (https://doi.org/10.5281/zenodo.3457213) and the **Leipzig-Berlin Declaration** (http://doi.org/10.5281/zenodo.3895208).

Main contributions of NFDI4Health will be on: ***Standardisation****:* (Co-)applicants of NFDI4Health have leading roles in relevant domain-specific (inter-)national standardisation bodies (ISO/CEN/DIN) and organizations (CDISC, HL7, GA4GH, OHDSI, IHE, LOINC, etc.), as well as meta-initiatives such as the European network EU-STANDS4PM and the Research Data Alliance (RDA) . ***Monitoring of data quality****:* Quality of data to be shared has to be closely monitored according to predefined criteria. Here, NFDI4Health will build on the DFG-funded project Standards and tools for data monitoring in complex epidemiological studies where a broad epidemiological community was involved under the lead of UM Greifswald. ***Federated data analysis***: NFDI4Health will provide experience in the development and implementation of federated data analysis infrastructures to all other NFDI consortia. ***Data management tools and data sharing models****:* NFDI4Health will develop tools, e.g., to create data management plans, processes and models for data sharing of sensitive person-related data, extending generic metadata standard such as DataCite and formulating requirements for archiving metadata and data sets. ***Data privacy, data protection laws and record linkage****:* Given the specific requirements with respect to data privacy, NFDI4Health will provide key knowledge and expertise regarding individual data protection and its implementation for the exchange of study data. NFDI4Health will develop solutions to enable shared use of person-related data in Germany and will give recommendations for necessary revisions of German law, in particular with respect to record linkage. ***Training and education:*** NFDI4Health will provide specific training in good practice of health data collection, data management, data access constraints and correct use/analysis of data. This does not only concern primary data collections but also the secondary use of existing databases and registries. Moreover, together with NFDI4BioDiversity, the city and state of Bremen and the U Bremen Research Alliance, NFDI4Health has started to establish a graduate education programme on research data management and data science.

Our expectations of participating in this NFDI conference:

- Information about planned and upcoming NFDI consortia to identify possible collaborations
- Networking with people from other consortia
- Clarification about formal establishment of NFDI as a whole and required contracts with NFDI bodies and within NFDI4Health
- Information about the NFDI Directorate
- Clarification about handling of cross-cutting topics and their implementation in NFDI

Members of the consortium (co-spokespersons)

Co-Spokesperson	Affiliated institution
Prof. Dr. Wolfgang Ahrens Deputy Scientific Director BIPS and Head of Department Epidemiological Methods and Etiological Research ahrens@leibniz-bips.de	Leibniz Institute for Prevention Research and Epidemiology – BIPS Achterstraße 30, 28359 Bremen
Prof. Dr. Iris Pigeot Scientific Director BIPS pigeot@leibniz-bips.de	Leibniz Institute for Prevention Research and Epidemiology – BIPS Achterstraße 30, 28359 Bremen
Prof. Dr. Hajo Zeeb Head of the Department of Prevention and Evaluation zeeb@leibniz-bips.de	Leibniz Institute for Prevention Research and Epidemiology – BIPS Achterstraße 30, 28359 Bremen
Prof. Dr. Fabian Prasser Head of the Division Medical Informatics fabian.prasser@charite.de	Charité – Universitätsmedizin Berlin, Berlin Institute of Health (Charité/BIH) Anna-Louisa-Karsch-Straße 2, 10178 Berlin
Prof. Dr. Sylvia Thun Director for eHealth and Interoperability sylvia.thun@bihealth.de	
Prof. Dr. Matthias Schulze Head of the Department Molecular Epidemiology mschulze@dife.de	German Institute of Human Nutrition Potsdam-Rehbruecke (DIfE) Arthur-Scheunert-Allee 114-116, 14558 Nuthetal
Dr. Oya Beyan Deputy Head of the Department Knowledge Pipelines beyan@fit.fraunhofer.de	Fraunhofer-Gesellschaft (Fraunhofer Institute for Applied Information Technology FIT) Hansastraße 27 c, 80686 Munich
Prof. Dr. Holger Fröhlich Head of the Division Data Science & AI, holger.froehlich@scai.fraunhofer.de	Fraunhofer-Gesellschaft (Fraunhofer Institute for Algorithms & Scientific Computing SCAI) Hansastraße 27 c, 80686 Munich
Prof. Dr.-Ing. Horst Hahn Institute Director of Fraunhofer MEVIS horst.hahn@mevis.fraunhofer.de	Fraunhofer-Gesellschaft (Fraunhofer Institute for Digital Medicine MEVIS) Hansastraße 27 c, 80686 Munich
Martin Golebiewski Research Associate Scientific Databases and Visualization martin.golebiewski@h-its.org	Heidelberg Institute for Theoretical Studies (HITS gGmbH) Schloss-Wolfsbrunnenweg 35, 69118 Heidelberg
PD Dr. Wolfgang Müller Scientific Director HITS wolfgang.mueller@h-its.org	
PD Dr. Sebastian Klammt Managing Director KKSN e.V. sebastian.klammt@kks-netzwerk.de	Network of Coordinating Centres for Clinical Trials – KKSN e.V. Alt-Moabit 96A, 10559 Berlin

Co-Spokesperson	**Affiliated institution**
Prof. Dr. Tobias Pischon Head of the Molecular Epidemiology Research Group tobias.pischon@mdc-berlin.de	Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC) Robert-Rössle-Straße 10, 13125 Berlin
Annette Polly Head of Department Research Services and Information Management Annette.polly@mri.bund.de	Max Rubner-Institut (MRI) Haid-und-Neu-Straße 9, 76131 Karlsruhe
Prof. Dr. Lothar H. Wieler President RKI president@rki.de	Robert Koch Institute (RKI) Nordufer 20, 13353 Berlin
PD Dr. med. Linus Grabenhenrich Head of Research Data GrabenhenrichL@rki.de	
Sebastian C. Semler, Executive Director of TMF sebastian.semler@tmf-ev.de	Technology, Methods, and Infrastructure for Networked Medical Research (TMF) Charlottenstraße 42/Dorotheenstraße, 10117 Berlin
Dr. Annette Pollex-Krüger Scientific Consultant annette.pollex-krueger@tmf-ev.de	
Prof. Dr. Ute Nöthlings Head of the Department Nutritional Epidemiology noethlings@uni-bonn.de	University of Bonn (Institute for Agricultural Science) Regina-Pacis-Weg 3, 53113 Bonn
Prof. Dr. Benedikt Buchner Institute Director IGMR bbuchner@uni-bremen.de	University of Bremen (Institute for Information, Health and Medical Law (IGMR)) Bibliothekstraße 1, 28359 Bremen
Ralf Depping Head of the Division Research and Publication Support depping@ub.uni-koeln.de	University of Cologne (University and City Library of Cologne (USB)) Albertus-Magnus-Platz, 50923 Cologne
Dr. Jens Dierkes Deputy Head of the Division Research and Publication Support dierkes@ub.uni-koeln.de	
Dr. Hubertus Neuhausen Institute Director USB direktor@ub.uni-koeln.de	
Dr. Harald Kusch Senior Scientist Medical Informatics harald.kusch@med.uni-goettingen.de	University of Göttingen, University Medical Center, Department of Medical Informatics Von-Siebold-Strasse 3,, 37075 Göttingen
Prof. Dr. Ulrich Sax Deputy Director Medical informatics ulrich.sax@med.uni-goettingen.de	

Co-Spokesperson	**Affiliated institution**
Prof. Dr. Carsten Oliver Schmidt Head of Division Quality in the Health Sciences carsten.schmidt@uni-greifswald.de	University Medicine Greifswald (Institute for Community Medicine), Department SHIP-KEF Fleischmann Str. 8, 17475 Greifswald
Prof. Dr.-Ing. Dagmar Waltemath Head of Division Medical Informatics dagmar.waltemath@uni-greifswald.de	
Dr. Oana Brosteanu Managing Director ZKSL oana.brosteanu@zks.uni-leipzig.de	University of Leipzig (Institute for Medical Informatics, Statistics and Epidemiology (IMISE); LIFE-Research Centre for Civilisation Diseases; Centre for Clinical Trials (ZKS)) Härtelstraße 16-18, 04107 Leipzig
Matthias Löbe Senior Scientist Medical Informatics matthias.loebe@imise.uni-leipzig.de	
Prof. Dr. Markus Löffler Institute Director IMISE markus.loeffler@imise.uni-leipzig.de	
Dr. Frank Meineke Senior Scientist Medical Informatics frank.meineke@imise.uni-leipzig.de	
Prof. Dr.-Ing. Toralf Kirsten Head of the Division Database Systems toralf.kirsten@hs-mittweida.de	University of Applied Sciences Mittweida Technikumplatz 17, 09648 Mittweida
Prof. Konrad Förstner Head of Department Data Science and Services foerstner@zbmed.de	ZB MED – Information Centre for Life Sciences Gleueler Straße 60, 50931 Cologne
Birte Lindstädt Head of Department Research Data Management lindstaedt@zbmed.de	