1 Binding letter of intent as advance notification of a full proposal

☒ Binding letter of intent (required as advance notification for proposals in 2021)

2 Formal details

- Planned name of the consortium:
  NFDI for Pre-clinical Drug Discovery and Chemical Biology

- Acronym of the planned consortium
  DeBioData

- Applicant institution
  Fraunhofer Institute for Translational Medicine and Pharmacology (ITMP)
  Theodor-Stern-Kai 7
  60590 Frankfurt a.M.
  Germany
  Head of the institution
  Prof. Dr. Dr. Gerd Geisslinger (geschäftsführend)
  Spokesperson
  Dr. Philip Gribbon
  philip.gribbon@itmp.fraunhofer.de
  Fraunhofer Institute for Translational Medicine and Pharmacology (ITMP)
  Schnackenburgallee 114
  22525 Hamburg

- Co-applicant institution
  Technical University of Munich
  Emil-Erlenmeyer-Forum 5
  85354 Freising
  Germany
  Head of the co-applicant institution
  Prof. Dr. Thomas F. Hofmann
  Co-spokesperson
  Dr. Stephanie Heinzlmeir
  stephanie.heinzlmeir@tum.de
  Technical University of Munich
  Emil-Erlenmeyer-Forum 5
  85354 Freising

- Co-applicant institution
  Helmholtz-Zentrum für Infektionsforschung GmbH
  Inhoffenstraße 7
  38124 Braunschweig
  Germany
  Head of the co-applicant institution
  Prof. Dr. Dirk Heinz

Co-spokesperson
Prof. Dr. Ursula Bilitewski
Ursula.Bilitewski@helmholtz-hzi.de
Helmholtz-Zentrum für Infektionsforschung GmbH
Inhoffenstraße 7
38124 Braunschweig

- Co-applicant institution
  EU-OPENSCREEN ERIC
  Campus Berlin-Buch
  Robert-Rössle-Str. 10
  13125 Berlin
  Germany
  Head of the co-applicant institution
  Dr. Wolfgang Fecke
  Co-spokesperson
  Dr. Katja Herzog
  katja.herzog@eu-openscreen.eu
  EU-OPENSCREEN ERIC
  Campus Berlin Buch
  Robert-Rössle-Str. 10
  13125 Berlin

- Co-applicant institution
  Universität Hamburg
  Bundesstraße 43,
  20146 Hamburg
  Germany
  Head of the co-applicant institution
  Univ.-Prof. Dr. Dr. h.c. Dieter Lenzen
  Co-spokesperson
  Prof. Dr. Matthias Rarey
  rarey@zbh.uni-hamburg.de
  Universität Hamburg
  Bundesstraße 43,
  20146 Hamburg
  Germany

Participants:

- Prof. Dr. Paul Czodrowski, TU Dortmund University
- Dr. Kamyar Hadian, HelmholtzZentrum München, Deutsches Forschungszentrum für Gesundheit und Umwelt (GmbH), Institut für Molekulare Toxikologie und Pharmakologie, Assay Development and Screening Platform
- Dr. Carsten Hopf, Center for Mass Spectrometry and Optical Spectroscopy (CeMOS)/ Hochschule Mannheim
- Prof. Dr. Manfred Jung, Institut für Pharmazeutische Wissenschaften, Albert-Ludwigs-Universität Freiburg

- Dr. Nicole Jung, and Prof. Dr. Stefan Bräse, Karlsruher Institut für Technologie (KIT), Institut für Biologische und Chemische Systeme (IBCS-FMS)
- Prof. Dr. Peter Kolb, Philipps-Universität Marburg
- Prof. Dr. Stefan Laufer, Eberhard-Karls-Universität Tübingen
- Dr. Alke Meents, Deutsches Elektronen-Synchrotron DESY
- Prof. Dr. Rolf Müller, Helmholtz Institut für Pharmazeutische Forschung Saarland
- Dr. Han Sun, Leibniz-Forschungsinstitut für Molekulare Pharmakologie im Forschungsverbund Berlin e.V.
- Prof. Dr. Yixin Zhang, Technische Universität Dresden, Center for Molecular and Cellular Bioengineering (CMCB), B CUBE, Center for Molecular Bioengineering

3 Objectives, work programme and research environment

- Research area of the proposed consortium (according to the DFG classification system: https://www.dfg.de/dfg_profil/gremien/fachkollegien/faecher/index.jsp)

Basic Research in Biology and Medicine	201
Microbiology, Virology and Immunology	204
Medicine	205
Computer Science	409

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

The consortia members are involved in the prosecution of research projects which lead to the creation of new data or insights into biological processes, specifically through the use of molecular and biological modulators. DeBioData activities cover the provision of infrastructure and services needed to maximise reuse of these and other community data sets. The key objectives aimed for by the consortium in the context of NFDI program are to enable a robust, efficient and qualified network of data infrastructures to extend knowledge about disease relevant biological mechanisms by facilitating the sharing of relevant pre-existing qualified data (with DOI, metadata, validated workflows) which have yet to be elevated to FAIR standards. We will align data providers, users, tool developers and relevant infrastructure platforms at German Research institutions, which share our common chemical biology and pre-clinical focussed goals. Moreover, the network-focussed approach implies that all German Research Institutes in any region can be associated with our activities, which will serve to maximise the impact of DeBioData across the scientific community. Users will be enabled to address their scientific questions making use of the accessibility of a larger data repository than previously, and, within the network be supported in the practical realisation of new analysis methods including machine-learning and artificial intelligence-based tools and workflows. This will allow users to generate, test and validate general prediction models and/or processes in their specific data domain within the broad field of chemical biology and pre-clinical discovery study results. The higher aggregation levels achievable in this way will pave the way to more precise models of human health and disease at the molecular, cellular, tissue and organismal levels. This will enhance the capacity of this community to identify, characterise and understand the influence of perturbagens such as compounds, drugs, chemical probes and biologicals which influence cell and tissue function. This information can be used both to understand disease states

and to identify starting points for the development of therapeutic interventions. Critical to the successful community-wide exploitation of these chemical probes resources is access to data sets and results covering their bioactivities, structure activity relationships, in-silico predicted properties, mechanistic and 'omics' effects on cells.

Objective 1: Promote scientific innovation in the life science and chemical biology communities by establishing an infrastructure for integrating, disseminating, searching, and re-analysing different types of relevant chemical biology and pre-clinical research data.

Objective 2: Engage with the chemical biology and pre-clinical research community to adopt FAIR data management, open science and data sharing principles. Provide training opportunities for researchers to learn and then implement FAIR methods

Objective 3: Identify and recommend standards for pre-clinical data and machine-readable meta data to support the FAIR principles for research data management in our community

Objective 4: Implement standards across existing standalone data repositories and data resources held by our community by means of a *DeBioData Schema* and local *MicroStores* based repositories which encapsulates the diverse and heterogeneous data types of modern chemical biology and pre-clinical research

Objective 5: Provide technical solutions which cater for the wide variety of purposes for which data were originally generated by data providers. Provide technical infrastructure and services solutions based on federated principles which maintain overall control by data providers, supporting intellectual property generation for time and competition-sensitive data whilst still realising efficient long-term reuse of data by the scientific community.

Objective 6: Develop synergies and implementable technical solutions with NFDI consortia working on "cross-community data" commonly used in pre-clinical R+D including imaging, genomics and structural data.

Objective 7: Demonstrate the utility and impact of the DeBioData concept using pre-selected use cases drawn from the consortia partners and in a second step high priority examples provided by the user community

Objective 8: Provide data originators/owners secure and compliant solutions based on an evolving data model (*DeBioData Schema*), a decentralized data warehouse allowing for storage (*DeBioData MicroStore*), query (*DeBioData QueryEngine*) and integration (*DeBioData Connection*) which will serve to elevate findability, access, interoperability and reuse of data

The Infrastructure will be delivered through 5 Task area.

Task area 1: Community engagement, development, training and support
Task area 2: FAIR policies and standards for Storage, Access, Interoperability and Integration
Task area 3: Infrastructure and technology development
Task area 4: Tools and Services
Task area 5: Coordination, Communication, Reporting

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

Our concept will be to strategically align data providers, users, tool developers and relevant infrastructure platforms at major German Research sites, which share our common research goals. Moreover, the network-focussed approach implies that all German Research Institutes in any region can be associated with our activities, which will serve to maximise the impact of DeBioData across the scientific community. Users will be able to address their scientific questions making use of a larger data repository than previously, and, within the network have access to a defined collection of results. In addition to the EU-OPENSCREEN network, TUM offers connections to national and international academic and industry partners of different disciplines such as the European Bioinformatics Institute (EBI), SAP and IBM. All of the partners have extremely strong international networks reflecting world-leading roles in data management related to their respective fields. Partner ITMP is strongly involved in the IMI-FAIRplus project as a work package lead as well as co-leading the cloudification work packages of the European Open Science Cloud (EOSC-) LIFE project, where EU-OPENSCREEN is also a partner.

In DeBioData, a database schema (*DeBioData Schema*) will be iteratively developed covering all scientific areas of DeBioData. The core entities (strong entities) like compounds, targets, diseases will be formally described with all attributes and representations. Primary and foreign keys will be defined for all entities to enable relationships between these core entities. A meta data model consistent over all data objects will be defined to handle partially conflicting multiple data points from different sources, measurement errors and data provenance. Based on the *DeBioData Schema* a software package *DeBioData MicroStore* for data providers will be created. This will be an easy-to-install local installation of an open source relational database system employing the *DeBioData Schema* involving a backend server and a customized web frontend for data providers. Our services will retain data on owner servers whilst making it accessible through maintained and sustainable web services or documented application programming interface (API) for data retrieval. The *DeBioData ConnectService* will use classical web server technology building on the *DeBioData Schema* and database. The *ConnectService* software framework will be instantiated for each external source. The *QueryEngine* is an open source implementation to access data from all available *Microstores* via their RESTful API's. It will implement a large variety of use cases and run on the consortium's web servers. Internal use cases and important data sources like PDB and ChEMBL will be realized as blueprint implementations.

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

Fraunhofer has extensive experiences in data management within multiple national and European programs (eg, Medicin Initiative, FhG Data Science program, EOSC-LIFE, and IMI FAIRplus). EU-OPENSCREEN (European Research Infrastructure for Chemical Biology) coordinates the data infrastructure activities of 21 screening and medicinal chemistry infrastructures in 8 European countries, including four German sites. Partner TUM is strongly engaged in the national and European proteomic community, exemplified by its leading role within the DKTK Proteomics platform and its participation in the European network project EPIC-XS. Partner UHH manages state of the art bioinformatics infrastructures including NERDD resource for drug discovery, the ProteinsPlus server for protein structures and SMARTS for visualization for SMARTS strings. The partner HZI is involved in extensive institutional data management activities specifically covering bacterial and viral pathogens, cellular imaging and toxicological data resources, respectively.

We have identified three initial “shared uses cases” to catalyse domain relevant cross-consortia collaboration with the NFDI consortia: **NFDI4Chem** and **NFDI4BIOIMAGE** and **InnoMatSafety** – The application of Medicinal Chemistry methodologies is a key part of small molecule drug discovery, pre-clinical research and chemical biology optimisation of compounds towards tools and leads. We will work with **NFDI4Chem** on screening data related standards and medicinal chemistry use cases. With **NFDI4BIOIMAGE** we will develop reproducible workflows and guidelines for handling high content imaging data with multi-feature annotation in pre-clinical drug development and plan to collaborate on the definition and use of image data formats and ontologies. With **InnoMatSafety** we will develop joint use-cases on the efficacy and toxicity of innovative materials used in the field of drug delivery and thereby evaluate interoperability of our data description standards. In addition to these specific co-operations we plan to actively collaborate with complementary NFDI consortia to facilitate data exchange and integration and promote scientific collaboration across scientific domains.

4 Cross-cutting topics

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

A substantial part of the overall DeBioData effort will be directed towards implementing FAIR approaches in chemical biology and pre-clinical research, by driving the successful application of ontologies and data / metadata standards. Therefore, the DeBioData consortium would stand to benefit from the work of cross-cutting NFDI initiatives working on these topics.

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

The DeBioData consortium would primarily act as a scientific domain-related infrastructure and would not directly contribute to cross-cutting topics. We aim to operate in the most collaborative manner possible and would be very willing to contribute to the efforts of other consortia, should the need and opportunity arise. We will interact with cross-cutting and newly operational consortia to find any areas of overlaps and collaboration opportunities. It is our aim to establish links to compute and storage-based infrastructure-based consortia and identify opportunities for cooperation.