Fraunhofer IME | Theodor-Stern-Kai 7 | 60590 Frankfurt a.M.

Deutsche Forschungsgemeinschaft
Kennedyallee 40
53175 Bonn
Germany

Fraunhofer
IME

Fraunhofer-Institut für Molekularbiologie
und Angewandte Oekologie IME

Institutsleitung
Prof. Dr. Dr. Gerd Geisslinger
(geschäftsführend)
Prof. Dr. Stefan Schillberg (kommissarisch)
Prof. Dr. Christoph Schäfers (kommissarisch)

Translationale Medizin
Leitung Prof. Dr. Dr. Gerd Geisslinger

Theodor-Stern-Kai 7
60590 Frankfurt am Main
Telefon +49 69 6301-7619 | Fax – 6301-7617
gerd.geisslinger@ime.fraunhofer.de
www.ime.fraunhofer.de

Frankfurt, 03.07.2019

Re: Non-binding letter of intent from the DeBioData consortium related to an anticipated submission of a full NFDI proposal in 2020.

To whom it may concern,

Please find below the letter of intent of the DeBioData consortium which is led by my colleague, Dr. Philip Gribbon.

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 Molecular Biology and Applied Ecology
Schnackenburgallee 114
22525 Hamburg
Germany

Head of the Institution

Prof. Dr. Dr. Gerd Geisslinger (geschäftsführend)

Spokesperson

Dr. Philip Gribbon
Fraunhofer-Institute for Molecular Biology and Applied Ecology, ScreeningPort
Schnackenburgallee 114
22525 Hamburg
Philip.Gribbon@ime.fraunhofer.de

Research area of the proposed consortium

Medicine

Objectives, work programme and research environment

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

DeBioData is a consortium assembled from German Life Science Institutions and associated Research Infrastructures working in the fields of pre-clinical research and human disease biology. The scope of the consortia's interests extends from basic scientific studies on molecular targets and their linkage to disease through to the analysis of late stage in-vivo proof-of-concept studies and their translation towards clinical investigations. The key objectives aimed for by the consortium in the context of NFDI program are to:

- stimulate and support basic scientific studies into pre-clinical research and associated human disease biologies, (including infectious diseases), and therefore help improve the currently poor translation between the lab-bench and the bedside
- 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 (Findable Accessible Interoperable and Reusable)
- whilst acknowledging the complexity of data types and standards, develop DeBioData combined with input from potential collaborating consortia, into a qualified infrastructure which links together a network of pre-existing indication, chemical, biological, 'omics and target-centric databases as well as any novel upcoming relevant resources, which collectively will form a unified resource for University, SME and large pharma industry based researchers
- work together with complementary NFDI infrastructures, tool providers, data originators and data users to define realisable standards for FAIR data in the disease target through to pre-clinical research domains by developing requirements and guidelines on the FAIRification processes

- deliver robust FAIR data methods for pre-clinical data-types, aligned to NFDI community developments and emerging European standards which individual German institutions and users can then adopt to integrate their own data resources into the DeBioData networked databases, tools and workflows
- provide data originators/owners and users secure and compliant solutions by means of novel blockchain-like solutions to assure data integrity
- allow scientific users to have a "one-stop shop" web-based service to search, find, collect and aggregate FAIR data important for realising their disease and pre-clinical projects

The standardization and harvesting of data on small molecules from reliable chemical-biological sources contribute to a better informed use of individual compounds and compound libraries in biological experiments. It will also help to properly assess the basis of biological responses and validate the hypotheses arising from the experimental results. AI and machine learning analyses will be facilitated through the provision of high quality imaging and phenotypic data sets based on "cell painting" readouts pursued with the EU-OPENSCREEN bioprofiling program. AI and machine learning will be enabled for global analysis across multiple data sets to reveal additional information and insights on chemotype-to-phenotype relationships with the potential to identify novel therapeutic targets and improved compounds

To achieve its vision of FAIRified pre-clinical research data as an accessible scientific resource to support the daily work of thousands of researchers, our concept will be to strategically align data providers, users, tool developers and relevant infrastructure platforms at major German Research sites and complementary NFDI infrastructures, which share our common research goals in the chemical biology and pre-clinical research domains. Moreover, our 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 in Germany and worldwide.

Planned direct contact with commercial service and instrument providers will enhance the consortium's capability to proactively influence shared solutions while adhering to the above vision. Users, finally, will be able to address their scientific questions making use of an interlinked data repository than previously, and, within the network have access to a defined collection of 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. The higher aggregation levels achievable in this way will pave the way to more precise models and enhance our capabilities to probe and understand the fundamental determinants of cell and tissue function and the deviations associated with disease and pre-disease states.

Together with identified interested German stakeholders and in anticipated cooperation with other NFDI projects, (identified below), the consortium will establish technologies which best can be adapted and implemented by single institutional users, thereby allow single researchers users to:

- be attracted to share (un)published data without losing their provenance and acknowledgement

- retain data on owner servers whilst data is also made accessible through maintained and sustainable web services or documented application programming interface (API) for data retrieval
- have access to common ontologies, standards and quality controls for their pre/clinical and data (e.g. detailed metadata and resource description language (RDF)
- use a central access process to access enabling the user to search and interoperate/integrate their workflows to expanded datasets which can be subsequently aggregated with complementary resources
- have access to the full spectrum of qualified target biology and pre-clinical research data from target genetics, medicinal chemistry, antibody design, bioinformatics workflows, "omics" databases, cellular phenotype data resources, to in-vivo efficacy, toxicological and ADME studies
- stimulate drug discovery by improving knowledge about biological mechanisms which will allow for development of novel therapeutic options
- predict toxic effects due to interfering with biological system and/or with ecosystems and identify targets for therapy or avoidance of toxicities
- drive translation potential by means of future integration and alignment with data infrastructures dealing with clinical data, a feature which will become increasingly in focus as GRDP-compliant data are routinely deployed in the future
- share/annotate data to go beyond uploading and subsequently forgetting about data and eventually allow researchers to integrate their data/tools with any other data/tools in the resource.

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

At the centre of the proposed DeBioData infrastructure is EU-OPENSCREEN (European Research Infrastructure for Chemical Biology), which coordinates the activities of 21 screening and medicinal chemistry infrastructures in 8 European countries, including four German sites. The RI will provide scientists access to chemical libraries, assay development and screening facilities, medicinal chemistry and informatics platforms as well as associated supporting facilities for protein production, cell line generation (e.g., 2D and 3D models and patient-derived cells), computational and structural biology as well as structure-based drug design. Data from these studies will be aggregated into the European Chemical Biology Database and the NFDI data services package. 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. In addition, TUM offers connections to national and international academic and industry partners of different disciplines such as the European Bioinformatics Institute (EBI), SAP and IBM.

FAIRification might be difficult to achieve for existing infrastructures, so in collaboration with complementary NFDI partners, we will establish a resource which is by design FAIR and integrates pre-clinical data types over the lifetime of the infrastructure. To achieve this, we will make use of new and emerging concepts (i.e. microservice architectures, cloud, blockchain, web 4.0/5.0) to enable non-experts to perform expert data manipulation and support hitherto challenging cross-resource analysis. The DeBioData Infrastructure will facilitate hypothesis generation to help elucidate the molecular determinants of diseases by providing interoperable FAIR data of assured quality which is suitable for using advanced algorithmic analyses of aggregated data sets. It will do this by providing documents, tools, workshops, e-resources, applications etc., to make data accessible and FAIR for all participants of the network.

The existing EU-OPENSCREEN infrastructure is located in Berlin and the DeBioData partners HZI and Fraunhofer IME represent 2 of the 4 German EU-OPENSCREEN Partner sites, selected by the BMBF. Partners TUM and MPI-CBG closely collaborate with EU-OPENSCREEN in i) Research and innovation action projects such as the H2020 project EU-OPENSCREEN-DRIVE on a European level and ii) with the Chemical Biology Interest group of DeCHEMA nationally. All of the partners have extremely strong international networks reflecting world-leading roles in their respective fields. Partner IME 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.

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

Initial discussions involving TC's or face-to-face meetings with six prospective collaborating consortiums have taken place in period since the DFG workshop event in Bonn, in order to identify concrete options for cross-consortia collaboration. While all discussions are still at the early stage and formal agreements are not yet in place, consensus positions are expected to be reached prior to a proposal submission in 2020, which would combine the concept of DeBioData with that of one or more other prospective NFDI consortia. Although, DeBioData remains open to wide range of collaboration options, the NFDI consortia for which the most compelling synergies could be identified were:

1) ***NFDI4Chem*** – 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 tool and leads. Discussions and exchanges of information between the respective consortium spokespersons have revealed multiple areas of common interest in the research data management infrastructures solutions needed to support working with compound physicochemical, target engagement, analytical, bioactivity, pharmacokinetic, toxicology and related datatypes. These synergies will be further explored in discussions between the 2 consortia going forward.

2) ***NFDI4Biological Imaging and Medical Photonics NFDI4BIMP***. Image based methods to quantify in-vitro pharmacological endpoints, as well as basic research studies linking targets, genotype and disease phenotypes by means of cellular imaging approaches are essential methods used by the DeBioData user community. DeBioData and NFDI4BIMP plan to collaborate on the definition and use of image data formats and ontologies shared across multiple NFDI consortia to facilitate image data exchange and integration and promote scientific collaboration.
3) ***BRIDGE4NFDI.*** As a generally applicable cross cutting consortia, working on FAIR methods, visualisation, metadata standards etc,, this consortia offered a complementary package of activities to that of DeBioData in relevant transversal topics. Potential collaboration opportunities will be further identified and developed in future discussions.

Cross-cutting topics

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 etc.. Therefore, the DeBioData consortium would stand to benefit from the work of crosscutting NFDI initiatives working on these topics. To explore this fully, our consortia has started initial discussions with ***BRIDGE4NFDI,*** in order to understand the real opportunities. Members of DeBioData are also actively engaged in related EU Projects such as IMI FAIRplus, EU-OPENSCREEN-DRIVE and EOSC-LIFE.

Topics your consortium could contribute to and how.

The DeBioData NFDI would primarily act as a scientific domain-related Infrastructure and would not directly contribute to cross-cutting topics. However, 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.