Leibniz Institute on Aging –
Fritz Lipmann Institute

1 Binding letter of intent as advance notification or non-binding letter of intent

☐	Binding letter of intent (required as advance notification for proposals in 2019)
X	Non-binding letter of intent (anticipated submission in 2020)
☐	Non-binding letter of intent (anticipated submission in 2021)

2 Formal details

- Planned name of the consortium

  Management of Primary Research Data

- Acronym of the planned consortium

  MOPED

- Applicant institution

  Leibniz Institute on Aging – Fritz Lipmann Institute (FLI)

  Prof. Dr. Alfred Nordheim

- Spokespersons

  Dr. Karol Szafranski (karol.szafranski@leibniz-fli.de)

  Prof. Dr. Steve Hoffmann (steve.hoffmann@leibniz-fli.de)

3 Objectives, work programme and research environment

- Research area of the proposed consortium

  Life Sciences (Biology, Medicine, Agriculture, Forestry and Veterinary Medicine)

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

  Primary research data is all such data that is collected directly from a primary data source. In Life Sciences as well as other disciplines primary data sources often are specialized measuring instruments producing electronic pictures, spectra or sequences of very different formats and structure. The quickly growing versatility of protocols and instruments make it increasingly harder to coherently annotate, organize, manage and archive the data produced by both older as well as

state-of-the-art measuring devices. While primary data usually only makes up a fraction of the entire data collection used in a research project, the characteristics of the primary data have profound ramifications for all subsequent analysis steps and experiments. Therefore, the management of primary research data is not only key to ensure findability, accessibility, interoperability and reusability of data (FAIR) but also to improve the reproducibility and verifiability of research results. While the level of standardization of formats and meta-data annotation is quite advanced for some types of instruments (e.g. high-throughput sequencing), other classes of primary data sources (e.g. gel cameras, optical devices, chromatographic devices) still lack satisfactory level of integration. We have begun to conceptualize strategies to facilitate a comprehensive primary research data management (RDM) with potential partners in Jena and beyond. In addition, we are evaluating the complementarity of our endeavour with other future NFDI consortia. The proposed program will rest on three pillars:

1. **Methodological:** Proposition of standards for the meta-data annotation and storage of primary research data from different instruments (e.g. instrument parametrization, sample information).

2. **Technical:** Development of strategies and tools for a convenient and immediate meta-data annotation as well as a data transfer from instruments (e.g. imaging cameras, chromatographic devices) to data storages with a maximum degree of automation (RDM at the source). Earliest-possible linkage of sample description meta-data with primary data.

3. **Educational:** Training of data stewards providing help for research groups to organize their (primary) research data according to the FAIR principles and to further develop standards and protocols in primary research data management.

In close collaboration with researchers, both technical and methodological developments are put into practice and are constantly adjusted by data stewards.

- Brief description of the proposed use of essential existing infrastructures, tools and services

The FLI distinguishes itself from other research institutions of similar size by its very wide variety of instruments for Life Science research. This includes a large diversity of high throughput sequencers (from PacBio to Illumina NovaSeq), flow cytometry analyzers (e.g. LSR Fortessa, FACS Verse), state-of-the-art mass spectrometers (e.g. Orbitrap Fusion Tribrid, LTQ Orbitrap XL ETD) or microscopes (e.g. Z1 Lightsheet, MD ImageXpress Micro Confocal) which are essential for a large number of wet-lab protocols. Our prospect cooperation partners will further add to this

diversity of instruments and workflows. Furthermore, the FLI operates a Life Science Computing Core Facility which is providing data management services to all research groups at the FLI. These services include an institute-wide electronic lab notebook as well as the management of a publication data archive intended to improve FAIR RDM at FLI. This set up allows us and our data managers to develop and deploy necessary technical solutions to allow a comprehensive RDM beginning right at the source. The research groups at the FLI, the partners of the future consortium and other researchers will be actively involved in all phases of the envisioned program. The FLI and its partners will make sure that all developments are reported, critically discussed and assessed with the users of the respective technologies. Complimentary to the efforts in developing appropriate primary data management protocols, the FLI and its partners will train data stewards helping to plan and organize the data management for specific research projects. Specifically, data stewards shall help to adjust the annotation, storage and archiving of primary and secondary research data to maximize FAIRness and to improve the swift reproducibility of research results. The experience of individual researchers and research groups shall than be re-integrated to refine primary data research management protocols and structures. The FLI and its prospect cooperation partners enjoy longstanding cooperations with producers and developers of measuring devices (e.g. Carl Zeiss Microscopy GmbH via the Aging Research Center). The consortium seeks to cooperate with those industrial partners to develop, implement and adjust the solutions for primary RDM.

- Interfaces to other proposed NFDI consortia

Naturally, the envisioned research proposal has interfaces with all proposed NFDI consortia interested in the management of Life Science data and medical data. This includes but is not limited to NFDI4Health, NFDI4Biodiversity or the overarching consortium NFDI4LifeUmbrella. Given the focus of our envisioned plan, i.e. to develop strategies for the management of the research data right at the source, MOPED has interfaces with a number of other projects outside the Life Science research area, e.g. NFDI4Chem. We have begun to make contact with other proposed NDFI consortia, e.g. NFDI4Health, to evaluate the opportunities for a collaboration or a participation.

4 Cross-cutting topics

- Identification of cross-cutting topics

First, the development of technical prerequisites to annotate, store and organize the research data right at its source might be of interest in all those consortia where data from a variety of instruments is being collected. The methodological objective of MOPED, i.e. to develop a comprehensive strategy for the primary RDM for the Life Sciences, critically involves a close interaction and collaboration with the experimental researchers. Questions concerning which metadata is required for a FAIR research data management or how much time a researcher is able to invest in annotating her/his data are most likely relevant for all of NFDI's consortia. Finally, also our educational focus, i.e. to train data stewards that assist in planning specific research projects, might be of interest for other consortia.

- Contribution to these cross-cutting topics

Within the course of the envisioned project, the FLI and its partners will develop software that allows the (preferably fully automated) registration and storage of primary research data. The underlying concepts for these solutions may be of interest for those consortia that incorporate data from highly diverse measuring instruments. Likewise, the proposition of annotation standards derived from the FAIR criteria and the rules of good scientific practice invite a multitude of cooperations with other consortium partners. Finally, also the practical education of data stewards to put the theoretical concepts and technical solutions into experimental research practice may be of common interest across the NFDI consortia.