# Letter of Intent for NFDI

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

*[Please indicate clearly whether your document is a binding letter of intent as advance notification or a non-binding letter of intent.]*

| | |
|---|---|
| ☐ | Binding letter of intent (required as advance notification for proposals in 2019) |
| ☑ | 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
  Facility for digitalized morphology of organisms

- Acronym of the planned consortium
  Di-Morph

- Applicant institution
  Universtität Heidelberg
  Rektor Prof. Dr. Dr. h.c. Bernhard Eitel

- Spokesperson
  Prof. Dr. Vincent Heuveline, vincent.heuveline@urz.uni-heidelberg.de, Universitätsrechenzentrum Heidelberg, Im Neuenheimer Feld 293, 69120 Heidelberg

3     **Objectives, work programme and research environment**

▪     Research area of the proposed consortium (according to the DFG classification system)

  ▪  www.dfg.de/download/pdf/dfg_im_profil/gremien/fachkollegien/ amtsperiode_2016_2019/fachsystematik_2016-2019_en_grafik.pdf

  The main topics that are addressed in terms of the research are due to the DFG classification are:

  -     2-21 (Biology)

  -     3-32 (Physics)

  -     3-33 (Mathematics)

  -     3-34 (Geosciences)

  -     4-44 (Computer Science, Systems and  Electrical Engineering)

- ▪ Concise summary of the planned consortium's main objectives and task areas
- ▪ [Your text, max. 2 pages]

With this NFDI, a data repository will be created, which will be successively developed and operated together with the disciplines as a virtual data facility. Essential elements of the structure, functions and services will be developed on top of the existing "Large Scale Data Facility" (LSDF2) and ressources of the private cloud infrastructure "heiCLOUD" in Heidelberg and further infrastructure and IT services (as described in the subsequent section).

As a first joint objective, the NFDI for Digitized Morphology of Organisms (Di-Morph) will be implemented. As a core element, the cooperation partners will create a smart repository for 3D morphological studies on organisms, which operates the Di-Morph Facility simultaneously as an interactive data storage and as a cooperative data analysis and communication platform. Data and services for the analysis are made available to the international scientific community. Research results will be quality assured and made accessible to the scientific and social public.

NFDI Di-Morph integrates four functionally linked hierarchical levels, using relevant algorithms to create and use digitized morphology: Building on (1) the digital acquisition and processing of experimental primary morphological data (biological image and metadata), (2) morphological secondary data are Volume data in the form of reconstructed and image-processed voxel data records - in addition to required metadata - generated. These are (3) analyzed by artificial intelligence (AI) procedures and used to generate tertiary data. (4) Bioinformatic approaches using AI are used to derive correlation data, hypotheses, and implications from linking morphological data to related genetic and / or environmental information.

The NFDI establishes a high compatibility of the facility (structure and services) both to the infrastructures of the primary data collection (for efficient data pipelines of large sample and data series) as well as the specifics of the disciplines involved (interdisciplinary basic concept). To demonstrate the potential of the Di-Morph Facility, concrete problems of the life sciences of high scientific and social relevance in the fields of environment and health will also be addressed.

The NFDI Di-Morph serves as the nucleus of the intelligent digitization of the morphology of organisms and, with the concept of the Smart Repository, develops and implements new and innovative strategies for the use of digitization as a key technology of the life sciences in Germany, Europe and beyond. From the beginning, Di-Morph aims beyond the immediate scientific impact on the applicability to further social problems in the environment and health.

- ▪ Brief description of the proposed use of existing infrastructures, tools and services that are essential in order to fulfil the planned consortium's objectives
- ▪ [Your text, max. 1 page]

| | |
|---|---|
| SDS@hd – Scientific Data Storage | Large scientific data in the multi-petabyte-scale can be stored on this online storage system; secure transfer protocols can be used to access the data; the usable capacity of ~8 PB is currently extended. |
| bwForCluster MLS&WISO | Intensive analysis steps can be done using this high-performance compute cluster. |
| bwVisu | This remote visualization service facilitates the use of massive CPU and GPU resources for interactive visualization and offers direct access to data stored at SDS@hd. |
| heiARCHIVE | Archive service for long-term preservation of research data on the level of bit-preservation with georedundant storage, online & offline storage with management of standardized metadata, cataloguing, and format conversion. |
| heiDATA | Heidelberg University's institutional repository for research data for all research fields. |
| heiCLOUD | Private cloud infrastructure (available also via DFN members) operated by Heidelberg University. |

**DFG**

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

The proposed NFDI consortia plan a strong cooperation with others NFDI consortia in the area of life sciences. Dedicated agreements with two Science Data Centers newly established in Baden-Württemberg are in planing. The goal of this closed interplay is to establish standards especially with respect to the setup related to Meta-Data, aggregation of data (interface OAI-PMH), persistent identifier and legal issues related to data curation.

## 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.
  - setup of the Meta-Data,
  - data curation
  - aggregation of data (interface OAI-PMH)
  - persistent identifier
  - legal issues
- Please indicate which of these cross-cutting topics your consortium could contribute to and how.
  - definition of an adequate setup of the Meta-Data,
  - new methods for exploratory data analysis
  - modelling and algorithms
  - AI based pattern recognition
  - Uncertainty quantification for data analysis
  - Aggregation of data