

Letter of Intent



NFDI4SD Small Disciplines

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

<input checked="" type="checkbox"/>	Binding letter of intent (advance notice of proposals due in 2020)
<input type="checkbox"/>	Non-binding letter of intent (expected submission date 2021)

2 Formal details

Planned name of the consortium: NFDI4SD Small Disciplines

Acronym of the planned consortium: NFDI4SD

Applicant institution

Humboldt-Universität zu Berlin, Prof. Dr.-Ing. Dr. Sabine Kunst,
Unter den Linden 6, 10099 Berlin

Spokespeople

Prof. Dr. Andreas Degkwitz, Universitätsbibliothek, Unter den Linden 6, 10099 Berlin;
email: andreas.degkwitz@ub.hu-berlin.de

Prof. Dr. Gerd Graßhoff, Institut für Philosophie, Humboldt-Universität zu Berlin;
email: gerd.graßhoff@hu-berlin.de

Co-applicant institutions

- Bayerische Staatsbibliothek, München
- CERN/Zenodo, Genf
- Johannes Gutenberg-Universität Mainz
- Kulturstiftung der Länder, Berlin
- Ludwig-Maximilians-Universität München
- Max-Planck-Institut für Wissenschaftsgeschichte Berlin
- Stiftung Preußischer Kulturbesitz, Berlin
- Technische Universität Berlin
- Zuse-Institut Berlin

Co-spokespeople

- Dr. Klaus Ceynowa, Bayerische Staatsbibliothek München; email: direktion@bsb-muenchen.de (angefragt)
- Prof. Dr. Markus Hilgert, Kulturstiftung der Länder; email: kontakt@kulturstiftung.de
- Prof. Dr. Anke Lüdeling, Humboldt-Universität zu Berlin; email: anke.luedeling@hu-berlin.de (angefragt)
- Prof. Dr. Volker Markl, Berlin Institute for Learning and Data (BIFOLD), Technische Universität Berlin; email: prof@dima.tu-berlin.de

- Prof. Dr. Klaus-Robert Müller, Berlin Institute for Learning and Data (BIFOLD), Technische Universität Berlin; email: klaus-robert.mueller@tu-berlin.de
- Prof. Dr. Kärin Nickelsen, Ludwig-Maximilians-Universität München; email: K.Nickelsen@lmu.de
- Prof. Dr. Eva Inés Oberfell, Juristische Fakultät, Humboldt-Universität zu Berlin; email: sekretariat.oberfell@rewi.hu-berlin.de (angefragt)
- Prof. Dr. Vivien Petras, Institut für Bibliotheks- und Informationswissenschaften, Humboldt-Universität zu Berlin; email: vivien.petras@bi.hu-berlin.de
- Prof. Dr. Jürgen Renn, Max-Planck-Institut für Wissenschaftsgeschichte Berlin; email: renn@mpiwg-berlin.mpg.de
- Prof. Dr. Dagmar Schäfer, Max-Planck-Institut für Wissenschaftsgeschichte Berlin; email: schaeferoffice@mpiwg-berlin.mpg.de
- Prof. Dr. Uwe Schmidt, Johannes-Gutenberg-Universität Mainz, Zentrum für Qualitätssicherung und -entwicklung, Arbeitsgemeinschaft Kleine Fächer; email: uwe.schmidt@zq.uni-mainz.de
- Prof. Dr. Christof Schütte, Zuse Institute Berlin (ZIB), Freie Universität Berlin; email: schuette@zib.de
- Prof. Dr. Tim Smith, CERN, Genf; email: tim.smith@cern.ch (angefragt)

Participants

- Malte Dreyer, Computer- und Medienservice, Humboldt-Universität zu Berlin; email: malte.dreyer@cms.hu-berlin.de
- Dr. Michael May, Head of Company Core Technology Data Analytics & AI Siemens, Siemens AG; email: May.michael@siemens.com
- Prof. Dr. Tonio Sebastian Richter, Ägyptologisches Seminar, Freie Universität Berlin; Berlin-Brandenburgische Akademie der Wissenschaften, email: sebastian.richter@fu-berlin.de

3 Objectives, work programme and research environment

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

101, 102, 103, 104, 105, 106, 107, 108, 109, 111, 112, 205, 207, 311, 314, 316, 409, 410

The more than 150 so-called small disciplines itemized by the University of Mainz's *Arbeitsstelle Kleine Fächer* are in many cases not listed as subject areas by the German Research Foundation (DFG). This means that, in some instances, several subjects fall within one group. For this reason and for the sake of simplification, we refer only to the larger groupings in this letter of intent.

Summary of the proposed consortium's main objectives and task areas

Digital transformation is fundamentally changing the way researchers work, particularly those working in the small disciplines. These scholars are typically involved in highly collaborative and global long-term projects which use innovative technology, but they often lack the research data and publication infrastructure normally provided by their home institutions. Collaborative projects require agile workflows for all user groups. Data aggregation, preparation, processing, analyses and publications are elements of modern scholarly research. The aim of the NFDI4SD consortium is to provide scholars with research-integrating data along with other scientific activities. By concentrating on the small disciplines, the consortium will be able to provide researchers with the micro-services that they require on a daily basis. The first objective will be to break up the traditional sequential research organization and weave research data, including the publication of the data, into an integrated research process, complemented by a moderated research model that weaves research data into the ongoing research workflow (the "cube").

Objective 1: *Both the use and production of research data will be tightly integrated into ongoing research projects. NFDI4SD services will be integrated into current research via an initial collaborative agreement between new research projects and the NFDI4SD. These agreements will be open to all disciplines, independent of their institutional classification as "small disciplines".*

Currently, the main desideratum of researchers working in the small disciplines is access to modern computational research data beyond the support of their home institutions and third-party funding bodies. As such, the NFDI4SD will be a novel institutional research partner in the scientific arena.

Objective 2: *The NFDI4SD will develop research data services that respond directly to the feedback of research projects. For many years now, the Arbeitsstelle Kleine Fächer has been recording the institutional settings of small disciplines. It enables direct communication between researchers, students, institutions and, in the future, the NFDI4SD.*

Modern concepts of computational philosophy of science will guide the software architecture of computational research data flows.

Objective 3: *The NFDI4SD will use and supplement best-practice tools and services within the European Open Science Cloud (EOSC).*

The NFDI4SD will be an active member of the German NFDI consortium and will strive to form collaborations with other suitable consortia. It will commit itself to the strategic plans of the EOSC Roadmap and seek membership in the newly created EOSC Association. The proven infrastructure of CERN's Zenodo service will provide OpenAIRE data publication, data harvesting and the implementation of FAIR data principles to implement the eight ambitions of Open Science.¹

Objective 4: *The signing of agreements with relevant stakeholders – libraries, archives and other content providers – on standards, application programming interfaces (APIs) and open access via computer networks.*

The NFDI4SD governing body will seek to make operational agreements with a number of content-providing institutions on the implementation and accessibility of APIs for the NFDI4SD's infrastructure hub. These agreements, using widely accepted standards, will enable researchers to access large sets of research data from a large number of content-holding institutions. It is expected that, within a short space of time, the NFDI4SD will be able to provide standardized API and interface modules. The large variety of small disciplines involved will ensure that special collections beyond the major stakeholders will be included in this integrated network of content providers.

Objective 5: *The NFDI4SD aims to maximize the visibility of the impact of research*

New generation metrics monitors will be implemented to monitor the use and impact of the NFDI4SD's services as part of the consortium's support of the research projects. Daily updated monitors and impact indicators will enable researchers to assess their collaborative global network as well as inform the NFDI4SD's governing body of hotspots of usage and of the need to steer users towards a particular course of action.

Objective 6: *Publication and citizen science*

All scientific output for general scholarly use will be regarded as published material. Such material goes beyond putting data on a file server: publications implement FAIR data principles and use review, curation and scholarly assessments. The NFDI4SD intends to establish new procedures and references for data publications in order to enhance to a significant degree the impact of research.

Task areas

The nature, benefits and characteristics of research data are surprisingly complex: data exist in many different media formats and contribute to the information value of the subjects of research; and researchers benefit from the rapid exchange of their publications and research findings. A Research

¹ Siehe [Open Science](#).

Management Plan will coordinate the use and interoperability of the NFDI4SD's services. Internal and external forms of communicating for research purposes, data storage, usage and revision as well as full-page archiving, including the final publication of research results and research data, will all form part of the NFDI4SD's services.

The workflow of research activities in the context of highly collaborative, agile scientific communities occurs at the interplay between data and theory, using – among other things – computational means. Such activities can neither be ordered as a linear sequence of tasks, nor as a research data life cycle. The organization of the workflow of research activities has been designed using Thomas Kuhn's metaphor of science as a puzzle-solving activity. We describe the pipeline of scientific processes as the transformation of input data via research pipes into designed outcomes. This sequence of operation can be compared to the rotations of a multidimensional research cube (solving Rubik's cube, for example). The cube's faces represent the content, data, skills, or means of the research activities. Each single step (aka cube's rotation leads to a new configuration).

This puzzle-solving metaphor facilitates the orchestration of the NFDI4SD's services into research activities by organizing and communicating the growing repertoire of the NFDI4SD's manifold of research data services. The cube describes the requirements, standards and quality of data, as well as their information, metadata and documentation, in order to assure the best usage of services. The NFDI4SD can confidently base its operation on a wide spectrum of standards, API norms, data formats as recommended by the European Open Science Cloud and other standard institutions. The NFDI4SD will rapidly develop a user-friendly service catalogue which will serve as a graphical user interface (GUI) for researchers, who will be able to choose their needed service, quickly apply it to their given data and research questions and obtain their intended results.

Task Area 1 (TA1): research fields

The aim of this large task area is to observe at close range the application of the NFDI4SD's services and then to recommend the development of additional services to satisfy the demands from the various disciplines. TA1 will be supported by a broad range of fields, assisted by coordinators.

TA2: services, machine learning, big data and data connectivity

The operational task area develops and secures the operation of the NFDI4SD's infrastructure. Using ZENODO utilizes Europe's largest and renowned research data publication platform and its comprehensive computer power.

TA3: collaboration tools, user interfaces and publications

Today's research environments are leading to a rising demand for cloud collaboration services. Direct exchange is being replaced by the publication of texts and data. User interfaces ensure that scholars and scientists can effectively use and control data without any special training. In addition, metadata as well as knowledge graphs, advanced catalogues and reference tools will enable users of the NFDI4SD's services to make the most of open scientific data.

TA4: standards, metadata and quality assessment

Data encoding, flow computers, computer interfaces, APIs and publications require the widest applicable standards, norms and metadata. Our principal investigators (PIs) are long-term members of the key standardization committees and so will ensure the interoperability of the research data over a long period of time.

TA5: institutions, governance, public outreach, media and literacy

The highly experienced and international scientific representatives responsible for this TA will ensure optimal information flow and the smooth undertaking of negotiations and agreements with research institutions worldwide. The organization of virtual conferences, newsletters, blogs, and hopefully of physical meetings in the future, should ensure productive and effective scientific communication and present the research results to the general public.

TA6: legal aspects

Intense scientific collaborative work and the exchange of information, including the production of publications, involve complex legal considerations. This area will be tasked with preparing the operating rules, information desks and consultation services. It will also actively shape the future legal landscape of the research activities of the digital transformation that we are currently undergoing.

Proposed use of existing infrastructures, tools and services

The *NFDI4SD Small Disciplines* integrates and offers institutional services of dozens of institutional partners:

Collaborative Research environments use cloud computing capacities implement and disseminates best available programmatic libraries in digital humanities and data analytical tools as a hub of world-wide open access resources. It will seek large scaling capacities from cloud and workstations to super-computing power, big data processing and long-term archiving. NFDI4SD will develop open access libraries for its cube computing pipelines and use interfaces.

- Data publishing, long-term archiving, OpenAIRE standards, harvesting and reliable storage will be provided by CERN, Zenodo according to FAIR principles.
- Computational container environments will be hosted by our computing centers: Humboldt's Computer and Media Service (CMS), Zuse Institute Berlin (ZIB), and computing centers from the German computing network.
- Containers and data will be redundantly mirrored.
- Frontend development in proven cloud environments and collaboration tools as gitlab.
- The most suitable and user chosen data services, programming packages from open source libraries will be installed and curated: NLP, OCR, machine learning libraries, div. scientific libraries as scikit-learn, language models, Image recognition, graph-processing, large data packages, 3D Processing. Intended is use driven demand of functional availability of open access software satisfying minimal formal and functional admission criteria.

Interfaces to other proposed NFDI4SD consortia: brief description of existing agreements for collaboration and/or plans for future collaboration

NFDI4SD Small Disciplines has explored collaborations with NFDI4Memory, NFDI4Objects, NFDI4culture and BRIDGE4NFDI4SD. We see large synergies strengthening the services for combined consortia.

4 Cross-cutting Topics

NFDI4SD implements a collaborative working environment with a unified GUI (Graphical User Interface) for many services and disciplines. These services scale from desktop modules to super-computing installations transforming complex research data objects using auxiliary components as input to an output research data object. These single process steps (NFDI4SD "cubes") map standardized input data to intended output objects. Cubes will be both developed by the NFDI4SD and submitted by a quickly growing user group who contribute their solutions to the community. An increasing catalogue of those components will promptly enhance the range of services. NFDI4SD cubes import, store, transform, connect and analyze research data for all departments in small disciplines. These services will be available to all other NFDI consortia as well.

- Center for data services using containers or Docker lakes, data services for databases, knowledge graphs, machine learning, OCR, text mining, language models, image recognition, 3D processing
- Publication units in cooperation with publishers and publication services provide maximum impact publication for research data
- Metadata harvesting beyond its core OpenAIRE infrastructure provided by Zenodo
- Legal advisory unit with machine learning diagnostic units for publication packages
- Simultaneous quality management and activity monitoring units indicate strength and demands for services