

Letter of Intent - NFDI4Earth

1 Binding letter of intent as advance notification

This is a binding letter of intent as required as advance notification for a full proposal submission in 2019.

2 Formal details

2.1 Planned name of the consortium

NFDI Consortium Earth System Science / NFDI Konsortium Erdsystemforschung

2.2 Acronym of the planned consortium

NFDI4Earth

2.3 Applicant

- Applicant institution
Technische Universität Dresden, 01062 Dresden
Rector Prof. Dr.-Ing. habil. DEng/Auckland Hans Müller-Steinhagen
- Spokesperson
Prof. Dr. Lars Bernard, lars.bernard@tu-dresden.de,
Faculty of Environmental Sciences

3 Objectives, work programme and research environment

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

34 plus related fields from 11, 21, 22, 23, 31, 32, 44, 45

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

NFDI4Earth primarily addresses digital needs of researchers in Earth System Sciences (ESS). ESS comprise a sizeable number of disciplines and communities with the overarching aim to understand the functioning of all subsystems of the Earth system and their interactions. A common characteristic of ESS data is their spatio-temporal context. Pressing global challenges (e.g. climate change, water scarcity, land-use change, environmental pollution, natural hazards) call for intensifying national and international interdisciplinary efforts. Linkages between different subdisciplines and scientific approaches can be substantially strengthened through collaborative research data management (RDM) and knowledge extraction, leading to better science. A multitude of observations and model data in very high spatial, temporal and thematic resolutions leads to rapidly increasing data volumes. Thus, describing and assessing Earth system processes, their dependencies, and their changes urgently require more efficient workflows and more powerful analytic frameworks. Today, various services to support RDM in ESS exist. However, they are scattered and heterogeneous as well as mostly project based, and are lacking long-term perspectives.

NFDI4Earth responds to this growing complexity of digital requirements within ESS. As a community effort, NFDI4Earth builds on a consortium bringing together scientists from universities, universities of applied science, research institutions and public authorities. NFDI4Earth moderates supply and demand, and stimulates innovation and cultural change in the collaborative and connected use of large, heterogeneous data in the ESS. NFDI4Earth supports open science by promoting the use of open data services and open software repositories, and by fostering the implementation of institutional data policies and data management plans that enable collaborative research and seamless data publication. NFDI4Earth will provide simple, efficient, open and unrestricted access to all relevant Earth system data, scientific data management and data analysis services guided by FAIR principles, supporting the whole research data life cycle and will thus enable cutting-edge research in ESS and beyond. NFDI4Earth targets the consolidation and harmonization of research data related services in ESS, in order to offer a user-driven, innovation-friendly as well as sustainable service environment to researchers. In a community process, we have structured NFDI4Earth in six major task areas:

1. NFDI4Earth2Coordinate covers the overall coordination of the NFDI4Earth consortium, including the financial management. Through establishing a coordination office, it supports the NFDI4Earth steering board in establishing a common governance, controlling the project progress, the reporting, the consolidation, and quality assurance of NFDI4Earth recommendations, reports and white papers. Further, it realizes the communication and exchange within the consortium (workshops, virtual collaboration platforms, etc.) as well as in the communication with the wider public (Media, Website, etc.).

2. NFDI4Earth2Connect tackles the interfacing to other NFDI consortia (see 3.4), the integration of NFDI4Earth into the joint development of the NFDI and related cross-cutting tasks (see 4) as well as the design of a long-term operation and organizational model for NFDI4Earth. NFDI4Earth has strong links into international initiatives and will ensure international embedding of NFDI4Earth into infrastructures and networks as for instance ENVRI, ICOS, EPOS, EOSC, ISO, OGC, WDS, RDA, and WMO.

3. NFDI4Earth2Create; A strong and active community of ESS researchers and data infrastructure providers is a key pillar for a successful and sustainable establishment of NFDI4Earth. The community building process is – as already practiced during the consortium’s initialization phase (see <https://www.nfdi4earth.de>) – organized as an open stakeholder activation and joint learning and collaboration process, open for new participants during all project phases. The needs, requirements and best practices of both, ESS researchers and data infrastructure providers, will be continuously consolidated into blueprints for RDM in ESS. Community building and evaluation measures will address universities, universities of applied

sciences as well as the various research institutions in ESS in order to establish the NFDI4Earth as a joint initiative. Measures on capacity building will span from supporting curricula in ESS related *data science* as parts of teaching at universities towards the provision of educational offerings for *ESS research data managers*. In collaboration with the scientific communities (e.g. DVGeo, GeoUnion), reputation mechanisms for ESS data and software publications will be established. This makes RDM a matter of course for all ESS researchers.

4. NFDI4Earth4All will implement measures to establish a one-stop source and (virtual) service desk, supporting access for all researchers to relevant ESS data (including data from public institutions) as well as guiding ESS researchers to appropriate repositories for their data publishing and archiving needs. The task will establish user-support structures for the creation of ESS data management plans in close collaboration with the overall NFDI development. The aim is to implement harmonized workflows for quality assured research data and software publications usable for all.

5. NFDI4Earth2Consolidate will achieve a common understanding and a set of agreed standards, frameworks, and rules as the basis for a consolidated NFDI4Earth implementation. Building on existing and widely accepted standards (ISO, OGC, W3C, etc.) and service design principles, a common set of ESS gold standards (for metadata, encodings, ontologies, service interfaces, etc.) will be identified and regularly reviewed. Further measures will establish common criteria frameworks to evaluate the qualities of ESS research data and service offerings and to support related qualification processes for the contributions to NFDI4Earth. This allows the consolidation of the ESS related RDM services and offerings and lays the foundation for an operational NFDI4Earth. The task further envisions the development of a virtual, cloud-based ESS collaboration environment: Following a user-driven and demand-driven approach, we intend to integrate the multiple disciplinary databases in a common framework and make them uniformly available for cross-disciplinary data analytics.

6. NFDI4Earth4Future will pave the way forward to develop novel innovative ways to explore research data in the ESS. A trend scouting measure monitors new IT methods and technological developments and assesses their potential for the NFDI4Earth. An incubator lab provides funding to stimulate the prototyping of innovative data management and data analysis methods (e.g. visual analytics, automated annotations, collaborated scientific data analytics, machine learning, data fusion). Showcases in and across different fields (e.g. water research, geophysics, oceanography, meteorology) will test existing and newly created services and thus pioneer new approaches in research.

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

The evolving NFDI4Earth consortium has started to screen the landscape of existing infrastructures, services and collaboration tools supporting ESS and to initially match these against user requirements in a community process. This initial survey resulted in a list of 108 platforms and tools, which are only partially overlapping with the repositories listed on <https://www.re3data.org/>. Many of the screened platforms are the result of international projects with German participation, underscoring the need for compliance with international efforts.

The survey showed that well developed infrastructures exist across subdisciplines in ESS. Examples range from very specific platforms e.g. geochemical exploration tools for meteorites to more general data hubs for mineralogy or sediment analytics, to generic platforms for in-situ, space, and model data in the terrestrial, marine and climate domains. ESS data usage increasingly requires data integration across domains. For instance, satellite remote sensing data need to be combined with in-situ data stemming from mobile or static monitoring platforms with the data being stored in different repositories.

Very generic data publication platforms like PANGAEA have a high level of maturity, visibility and usage. These platforms contribute with their experience to realize consistent research data publications and archiving within NFDI4Earth. However, data integration, collaboration or analytic frameworks are currently not on offer there. The NFDI4Earth aims to close this gap by allowing the integration of data streams across different branches of ESS in a coherent way. The NFDI4Earth can capitalize on a series of initial consolidation efforts that share this vision, e.g. the Helmholtz Research Field "Earth and Environment" has started to create a DataHub across compartmental boundaries (Atmosphere, Marine, Terra). Another example exists at DKRZ as a well-established node in the international Earth System Grid Federation. Their hosting of Earth system model data across subdomains forms the basis for direct coupling to an observational database as provided by the German Weather Forecast (DWD) and future climate change assessments. Further initiatives have developed near real-time data services, large-scale sensor data management, online access, and metadata management.

While the majority of existing services are mainly dedicated to data search and download, the NFDI4Earth aims to connect to international efforts like the climate data store (ECMWF), the European Space Agency (ESA) data cubes, and the European Open Science Cloud (EOSC) to develop novel avenues to make scalable data analytics available to a wide user community in the cloud. This is why the NFDI4Earth also advances data integration and data analytics. In this context, the NFDI4Earth also represents a number of the German centers for scientific computing. Together we want to develop the next generation of services and tools that enable

Earth system scientists to explore a wide range of data interactively allowing for automatized data and metadata flows from sensors and models to data analytics and data and software publications.

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

NFDI4Earth has already established close contacts with Astro-NFDI, KonsortSWD, NFDI4Agri, NFDI4BioDiversity, NFDI4Ing, and NFDI4Objects. The consortia aim to develop a common roadmap towards the NFDI and will jointly elaborate on the cross-cutting topics (see 4). It has been agreed with these consortia that common, interoperable interfaces will be established to allow for seamlessly sharing ESS related data (e.g. on topography, geology, hydrography, climatology etc.) as indispensable elements for interdisciplinary research related to global challenges. NFDI4Earth provides its capacities and services for acquiring, sharing and processing a wide range of spatio-temporal data, as well as its competences and rich experiences in the development and consolidation of international data and metadata standards in the ESS domains to the other NFDI consortia. Joint work on establishing common standards for metadata and data will, for example, support integration of NFDI4Earth data offerings with data from biodiversity research, agricultural sciences, socioeconomic studies, archeology, astronomy and environmental engineering.

Moreover, for the joint development of the NFDI a set of NFDI commons (e.g. core standards, scientific cloud solutions, creating scientific data literacy, need for cultural change) is currently being elaborated together with NFDI4BioDiversity and shall help to establish the core pillars for the NFDI development.

4 Cross-cutting topics

4.1 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.

Several crosscutting issues have been discussed with other NFDI consortia. Common issues identified comprise

- NFDI governance model and sustainability, operational, cost-covering and legal models for the coordination bodies/offices of the NFDI and the NFDI consortia
- NFDI harmonization and coordination
 - on legal aspects (licensing, intellectual property rights and data protection) and ethical aspects of sharing research data and research software

- on terminologies (vocabularies and ontologies, reference systems, code lists); common data and metadata standards and encodings (e.g. NFDI core metadata) and unique identifier systems
- on data and service quality criteria, evaluation and qualification criteria, qualification and/or certification processes for NFDI service offerings and infrastructures, etc.
- International embedding of the NFDI consortia, safeguarding NFDI needs and requirements in European/international developments of research data infrastructures and in international standardization bodies (W3C, ISO, etc.)
- Establishing NFDI common shared services (e.g. computing services and scientific clouds, collaborative working environments, authentication and access mechanisms, registries, long-term archiving etc.)
- Establishing coherent NFDI user-support, linking to existing and emerging research data help desks and support units in the various research institutions and universities
- NFDI capacity building and education with activities towards research data literacy, capacity building on RDM at all levels, establishing professional RDM education and careers
- Stimulating a cultural change of ESS data users and providers towards FAIR and open research data, establishing scientific reputation for research data activities and engagement in scientific software developments, installing common conferences, graduate schools and research projects related to innovative aspects in RDM and data analytics.

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

- As expected from all NFDI consortia, NFDI4Earth will actively contribute to the NFDI governance model and sustainability models. The NFDI4Earth consortium did not yet establish as a legal entity; thus we also hope to benefit from common approaches and solutions.
- Several of the NFDI4Earth co-applicants and participants operate research infrastructures of various sizes and with different operation models. Building on this breadth of practices and experiences and meanwhile fully acknowledging the need for linking and consolidation of the existing plethora of offered services and tools the NFDI4Earth co-applicants and partners have a vital interest in developing sustainable NFDI operation models.

- NFDI4Earth builds on long standing active participations in various international networks. This applies to EES related initiatives as ENVRI-FAIR, IPCC, WMO, EOSC, and WDS, or the engagement in organizations as OGC to ensure the establishment of international standards or in the implementation of EU Directives as INSPIRE, CAFE or WFD. It also includes acting in cross-cutting initiatives as RDA or the EOSC-Hub. NFDI can capitalize these networks, engagements and experiences for the NFDI international embedding.
- NFDI4Earth will actively participate in the NFDI harmonization efforts. Several NFDI4Earth participants build on widespread and valuable experiences in contributing to international standards and agreements, most prominently in the area of metadata standards and data encodings for spatial data. NFDI4Earth profits from a well-established culture and skills concerning community software developments, which will be contributed into the NFDI.
- NFDI4Earth will serve the NFDI as knowledge hub for dealing with earth related spatio-temporal data – which serve as reference frame, search and visualization environment for many different disciplines and applications – and the associated data management and data analysis methods. NFDI4Earth plans common, cross-consortium showcases with other NFDI consortia (e.g. NFDI4BioDiversity, KonsortSWD) to evaluate the effectiveness of the NFDI and demonstrate the NFDI benefits. This will bring out a new quality in interdisciplinary research.
- In cooperation with NFDI4BioDiversity and further NFDI consortia NFDI4Earth plans to support or establish schools and academies for all scientific career levels on innovative aspects in ESS RDM and in ESS data science.
- NFDI4Earth will provide input to education and capacity building modules related to earth system data management and analysis.
- A number of co-applicant and partner institutions within NFDI4Earth operate HPC and storage infrastructures, which will contribute to the common shared services of the NFDI in general. The NFDI4Earth and the NFDI4Biodiversity community share the vision of a Common Scientific Cloud and the NFDI4Earth consortium can offer to actively co-design and support its stepwise implementation.