

1st NFDI Conference: Guidelines for

Extended Abstract Submission

Attending the NFDI conference requires submission of an extended abstract. The extended abstract will assist with preparation for the conference with respect to networking and discussions among participants, the NFDI expert committee and the DFG Head Office. The content is not binding, but should describe the current situation and needs from the perspective of the respective scientific communities.

Please note the following general preliminary remarks:

- The information set out in the extended abstracts is **non-binding** with regard to any later proposal and review, will **not be subject to any review process** and should reflect the **current status of discussions and planning**.
- The extended abstracts **serve the sole purpose of preparation for the conference** with regard to discussions and participant networking as well as organisation on the part of the DFG Head Office.
- The extended abstracts will only be made available to participants in the NFDI conference and the DFG Head Office.
- The description of the “2. Subject-specific and infrastructural focus of the planned consortium” must not exceed a maximum of six pages. The document must be structured according to the template provided (this rtf file, Arial 11).
- Extended abstracts must be written in English.
- Please send your extended abstract/this rtf file to the following email address:
nfdi@dfg.de
- **DEADLINE: March 29, 2019**

Please address the following aspects in your extended abstract:

1. Formal details

Planned title of the consortium
NFDI Konsortium Erdsystemforschung

Acronym of the planned consortium
NFDI4Earth

Lead institution or facility
Not yet decided, in process

Name and work address of a contact person (including email address and institutional affiliation)
Roland Bertelmann, roland.bertelmann@gfz-potsdam.de,

Helmholtz-Zentrum Potsdam
Deutsches GeoForschungsZentrum GFZ
Stiftung des öff. Rechts Land Brandenburg
Telegrafenberg, 14473 Potsdam

Members of the planned consortium (including institutional affiliation, without address)

Not yet decided, in process. Part of the discussion process are universities, all Leibniz Institutes in ESS, all Helmholtz Centres in Earth & Environment, relevant Max Planck Institutes, Alliances like ABC/J, Geo.X, German Marine Research Alliance, governmental institutions like BGR, BKG

Participants in the NFDI conference (names, institutional affiliation and email address; max. 3 persons)

Prof. M. Sester, Monika.Sester@ikg.uni-hannover.de, Universität Hannover

Prof. L. Bernard, lars.bernard@tu-dresden.de, Technische Universität Dresden

R. Bertelmann, roland.bertelmann@gfz-potsdam.de, GFZ German Research Centre for Geosciences

Research area of the planned consortium (research area according to the DFG classification system [not subject areas]: http://www.dfg.de/download/pdf/dfg_im_profil/gremien/fachkollegien/amtsperiode_2016_2019/fachsystematik_2016-2019_en_grafik.pdf
34 plus related fields from 21, 23, 31, 32, 44

Participating research institutions (without address)

Not yet formalized, part of the discussion process are universities, all Leibniz Institutes in ESS, all Helmholtz Centres in the Research Field Earth & Environment, relevant Max Planck Institutes, Alliances like ABC/J, Geo.X, German Marine Research Alliance, governmental institutions like BGR, BKG, associations like DVGeo

Participating infrastructure facilities and/or potential information service providers (without address)

Not yet formalized, but part of the discussion process are relevant infrastructures provided by governmental institutions, Helmholtz centres, Leibniz institutions, DKRZ, PANGAEA

Planned proposal submission date

2019

Overview diagram or organisational chart for the planned consortium

The planned Governance of NFDI4Earth addresses:

1. Establishment of a management structure that is able to first identify and subsequently respond dynamically, effectively, efficiently and fast to needs of ESS for research data management.
2. Building up a national research data infrastructure in an open and flexible process to adjust continuously to the needs of a large community. Clear funding rules and flexible Funding schemes should support i) active involvement of as many stakeholders as necessary and possible through funding of activities in single working packages, even for stakeholders that are not members of one of the applicant institutions and ii) flexible short-term funding for newly arising issues.

Continuous monitoring and adjustment of the process as well as coordination of the various activities is realized by a NFDI4Earth Steering Group. It consists of representatives of the applicant institutions, representatives of the working groups, and representatives of scientific and other stakeholders relevant to the scope NFDI4Earth. The Steering Group is seconded by a Scientific Advisory Board of high-level experts. Governance rules will define the establishment of the Steering Group and the interplay within the various NFDI4Earth Working and Stakeholder Groups. Preliminary chart:

2. Subject-specific and infrastructural focus of the planned consortium

Key questions/objectives of the consortium

NFDI4Earth will provide simple, efficient, open and unrestricted access to all relevant earth system data and data services guided by FAIR principles, enabling cutting-edge research in Earth System Sciences (ESS) and beyond. This objective will be achieved on a consortia basis anchored in the width of the scientific community.

NFDI4Earth uses its label “Earth” as an abbreviation for Earth System Sciences (ESS). A typical characteristic of ESS data is their spatio-temporal context with geo-references. ESS

itself is an umbrella for a number of disciplines and corresponding communities that come together to answer questions relating to our dynamic living environment. ESS combines the historically evolved individual disciplines to form a quantitatively working and highly interdisciplinary natural science, which pursues the integrated systems concept. Its methodology is derived from the disciplines of chemistry, physics and biology in order to capture the complex interwoven, coupled and dynamic processes in high precision and resolution to depict them in elaborate models. ESS can be thought of as a tree trunk that ramifies into different disciplines, each of the disciplines (e.g. geology, geodesy, hydrology, meteorology, oceanography, etc.) ramifying further in sub-disciplines and foundation sciences. The disciplines and communities are neither strictly delimited from other scientific disciplines or scientific communities nor from environmental agencies and authorities.

Pressing Global Challenges (climate change, water scarcity, hazards) push ESS even more for national and international interdisciplinary approaches in a strong linkage of fundamental and applied sciences. NFDI4Earth thus addresses the growing complex digital requirements of the individual disciplines and sub-disciplines within ESS. In addition, complex digital requirements arise from the increasing demand for interdisciplinary scientific approaches we need to answer and newly emerging data analyzation methods.

Growing availability of a multitude of measurements and simulations in very high spatial, temporal and thematic resolution result in rising complexity and in quickly growing amounts of data describing the earth system. NFDI4Earth has the mission to provide - as elementary basis for cutting-edge research - scientists in Germany and worldwide with simple, open, efficient, limitless and timely services to support both: (1) access and query the available data and (2) qualifying and publishing research results.

Special attention is therefore paid to (1) the development of a common concept for research data access and publishing and its implementation as part of the NFDI, (2) the technical development and establishment of usable, subject-specific research data workflows and trustworthy data services, (3) a process to guarantee consolidation and sustainable maintenance of these data services building on existing offers and (4) capacity building related to research data management and data science for ESS.

NFDI4Earth understands its initial as well as the future activities in community building as a 'community learning process' largely open for new 'players' (also during the project phase/s) and having to evaluate given examples and integrate experience into the current enterprise from the very beginning.

- **Known needs/current status of research data management in the relevant discipline/subject-specific relevance of the planned consortium:**

- o **From a research perspective**

Multiple fundamental advances in the ESS depend first and foremost on unrestricted access to data from different sources, which can be easily combined, analyzed and shared within the research community. Common standards and semantics as well as operational processes and services for data curation, publishing, hosting and archiving have been established by various institutions and projects. However, a consistent application of the FAIR principles considering not only data collections but also related processes and methods for analysis is not easily achievable. Further, the related services need to be supplemented by principles for sustainability, which is in general not given yet.

Approaches differ widely and ESS is far from supporting easy seamless integration of different data streams. This applies to all four aspects of FAIR. For example, the diversity of data sources makes it difficult for researchers to locate the best offer. Beyond the lack of a consistent implementation of FAIR principles, researchers nowadays face several infrastructural challenges, e.g.:

- The data volume in ESS is rapidly growing and often too large to download data for

local processing, approaches to support distributed and cooperative processing are still in their infancies.

- Co-interpretation of multiple heterogeneous and decentralized data streams is currently not possible in most of the given platforms, in particular with respect to cross-disciplinary applications.
- Several researchers (esp. at universities) lack common guidelines and guidance and ultimately infrastructures to realize research data management.
- Demand for education and training: (a) Research data handling along all phases of the data life cycle (b) Thematic research towards ESS-related Data Science, which emphasizes more and specific training in mathematics and computer science.

Hence, researchers need platforms that integrate data access and publishing options across archives, and link data to easy-to-use scientific computing facilities.

o In terms of available information providers and services

Multiple generic and specialized platforms have been developed in ESS and are widely used, like

- Data collections of in-situ platforms
- Remote sensing platforms
- Gridded Earth observation sharing platforms
- Model-data (sharing, evaluation, intercomparison) platforms
- Geodata Portals

Most platforms have the potential to (partially) implement the FAIR principles. Some are more advanced than others, e.g. in terms of adding DOIs and advanced findability.

Nonetheless, there are clear deficits as intelligent networking of these platforms for interdisciplinary questions is concerned. Also sustainable availability and preservation of methods and processes, e.g. software, resulting in lack of support for traceability, plausibility and reproducibility is not guaranteed.

• Summary of the planned research data infrastructure that is specifically intended to address the needs of research users in their respective work processes

In the long run, NFDI4Earth infrastructures shall allow researchers to find and publish data across domains and sub-disciplines, evaluate their quality and ingest them into their scientific workflow. Ideally, this should be possible based on virtual platforms that enable to connect data streams and at the same time allow to inspect data and implement, deploy and perform complex analytics. Further, the outcomes must be shareable according to the FAIR principles and workflows should be sharable and traceable. To implement these goals, (a) a series of powerful data hubs for intelligent querying, accessing, publishing and processing research products will be made available and (b) NFDI4Earth data hubs and tools enable data explorations with machine learning, visual analytics and other tools. Such a development and consolidation process needs to be steered by a continuous gathering and reflecting of user needs. Thus, active and holistic involvement of the scientific user community is core – but still far from current practice and requires the design and implementation of new mechanisms.

NFDI4Earth builds on a phased approach. The first phase of NFDI4Earth will pave the way for the main infrastructural goals, by

- identifying the demands of/for digital changes in the German ESS community in a widely structured community consultation process.
- establishing a set of common principles, rules and standards to serve as guidance for ESS researchers as well as a qualification frame for NFDI4Earth services and platforms
- establishing experimental prototype platforms, operating on distributed resources and serving as blueprints for following development and capacities. These platforms shall
 - a. support a series of specific use cases coming from different sub-disciplines in Earth system sciences.

- b. develop sustainability concepts implying reusable and well-documented underlying processes and software technologies.

In the midterm, all NFDI4Earth infrastructures shall allow researchers to share their data compliant with the FAIR principles even when these data streams are large. This implies, that (a) everyone has access to, a good understanding of, and contribute to specialized and generic platforms for sharing data according to the FAIR principles, (b) NFDI4Earth provides SMART guidelines for publishing data and simple processes in Earth sciences and helps individual researchers and institutions in Germany and worldwide to achieve them and (c) Researchers are duly acknowledged for data sharing.

As a result, NFDI4Earth infrastructures are competitive to, yet compatible with, other international initiatives and Germany is at the forefront of Earth system data research infrastructures.

- **Description**

- o **of data types**

NFDI4Earth has to deal with a broad variety of data types due to the range of subjects and subdisciplines in ESS. Data stems from observations, simulations and analysis in different combinations, forms and formats (numerical, textual, and graphical). Many different open and standardized as well as proprietary data formats as well service interfaces are used.

Relevant standardization stems for instance from the World Wide Web Consortium (W3C), the Open Geospatial Consortium (OGC), ISO/TC 211 as well as from legal framework as the EU INSPIRE directive. Just to name some examples: XML data are widely used for domain specific schemata (e.g. GeoSciML, WaterML, BorholeML etc.), data formats as netCDF/ HDF5 and GeoTIFF are most commonly used to exchange meteorological, hydrological and remote sensing data. Topographic data is often exchanged using GML, GeoJSON or ESRI's Shape format. Several approaches address common vocabularies/ ontologies, but still it lacks in providing (semantic) data mapping and transformation paths and data integration is still very tedious.

- o **of underlying data processing / data analysis methodologies**

Due to the thematic range of NFDI4Earth, a large number of methods is being used and ranges from numerical and statistical modelling, over machine learning to semi-automated and human interpretation. Core commonality of EES data (and the processes using it) is the space-time reference, which can be found on different scales. Examples span from the analysis of data in earth system data labs or large model comparison studies towards field studies and stakeholder interviews in the field of climate change research. As an addition and exemplary, often complex digital processes are established, which mostly require specific scientific expertise, as for example in the acquisition and analysis of observation data from satellite measurements. ESS often faces a data paradox: It has to adapt dealing with on the hand mass data (e.g. from remote sensing) or big data (e.g. social media) with, on the other hand data scarcity, i.e. spatially, temporally and thematically inhomogeneous distribution of observations and information.

NFDI4Earth will make an important contribution to mutual knowledge and learning about methods and processes in this scientific field and can in particular strengthen interdisciplinary research in ESS.

- **Planned implementation of the FAIR principles and information about any existing policies or guidelines in the relevant discipline**

NFDI4Earth implements FAIR principles on the base of the current state of affairs.

NFDI4Earth can hereby build especially on two initiatives on the European and the international level. Partners within NFDI4Earth are already engaged in these initiatives:

Support and rollout for Germany of "Enabling FAIR data in the earth, space and environmental sciences" (<https://copdess.org/enabling-fair-data-project/>). This means that repositories within NFDI4Earth will strive to:

- Ensure that research outputs (e.g., data, software, technology, and physical samples) curated by repositories are open and FAIR, have essential documentation, and include human-readable and machine-readable metadata (e.g., on landing pages) in standard formats that are exposed and publicly discoverable.
- Ingest and expose data to promote interoperability and reuse.
- Ensure that unique, persistent identifiers are used for authors (e.g., ORCID), research objects (e.g., Digital Object Identifier), and physical samples (e.g., IGSN).
- Create associations among the research outputs that they manage and other related entities.
- Ensure that data and software have licenses that are as open as possible, and as protected as necessary.
- Gain third-party validation of trustworthy and sustainable practices and capabilities.

Furthermore, the ENVRI community has set up the project “ENVironmental Research Infrastructures building Fair services Accessible for society, Innovation and Research” (<http://envri.eu/envri-fair/>). Members of NFDI4Earth are involved in both approaches and can therefore ensure an implementation within NFDI4Earth (e.g. DKRZ, FZJ, GFZ, PANGAEA, TIB). Following this path NFDI4Earth is in a perfect role to fit into programmes like GO FAIR. As education in research data management and data sciences will be an important part of NFDI4Earth the integration of the FAIR principles in all initiatives on Earth data literacy will be crucial.

- **Planned measures for user participation and involvement**

Especially in the field of ESS an inspiring and fruitful culture of exchange and cooperation has existed for a long time, which manifests itself in various forms of networking in professional societies, multidisciplinary research communities and international research alliances as well as participation in national and international research infrastructures. The consortium will be an integral partner in these networks and generates substantial added value.

Due to the large heterogeneity of communities within the Earth System Sciences, there is a plethora of existing initiatives and infrastructures. Thus, one of the tasks of the NFDI4Earth consortium in the first phase will be to perform a systematic survey of existing infrastructures and initiatives in ESS.

NFDI4Earth recognizes and respects the diversity of the ESS and makes different efforts to bring all relevant disciplines and sub/communities as well as actors together from the early beginning - in an open bottom up oriented process. We recognize these activities mainly as a community learning process, having to evaluate given examples and integrate experience into the current enterprise. This shall be organized by interest and working groups (as already established in RDA). In parallel, NFDI4Earth can also build on the strong support of institutions and make use of the broad range of scientific societies in the field and collaborates with the overarching organizations DVGeo and GeoUnion. Although awareness of the need and benefits of an open data culture in Earth System Science is on the rise, there is still a lack of knowledge and data literacy. Thus, research data management skills need to be developed on all levels from bachelor students to senior scientists. With the help of university partners, scientific societies, and well-established regional networks, NFDI4Earth will foster the discussion on appropriate peer-to-peer teaching and training formats as well as on a change in acknowledgement of data curation and publication. Envisioned tools for this can be e.g. trainings for scientists, modules of university curricula, raising awareness for research data management, implement data preparation as a self-evident part of research and academic teaching at all career levels and across all faculty positions, strengthen the academic value for data publications, data citations etc. in appointment and recruitment processes.

- **Existing and intended degree of networking of the planned consortium**
 - o **nationally (in particular with other, potential future consortia or existing**

state-level initiatives)

As Earth System Sciences have many interfaces with other fields of science, engineering and also social sciences, future links and collaborations are envisaged. Thus, rather than a completely stand-alone initiative, NFDI4Earth will be part of a network linking various disciplines. The NFDI4Earth initiative has already relations to NFDI4BioDiversity, potential links lead to NFDI4Chemie and NFDI4Health and NFDI4Microbiome. On the other hand, there will be thematic links to NFDI4agrar and NFDI4ING and the FAIRmat initiative in material sciences. The latter might touch upon the field of solid earth geophysics in NFDI4Earth, as well as the Astro-NFDI. Beyond that, there are connections to KonsortSWD on the level of societal impact of ESS, i.e. climate change research. With respect to consortia like NFDI4Memory, NFDI4Culture and NFDI4Text+, a common interest is in historical maps, scientific literature and data.

Networks like Geo.X and Geoverbund ABC/J or associations like the WaterScienceAlliance are already connected to NFDI4Earth and bring in their experience in successful networking universities and research institutions as well as stakeholders from outside academia. On a state level NFDI4Earth will cooperate with the relevant scholarly societies and can build on preliminary work e.g. of DFG Specialised Information Services like FID Geo.

Of special interest is networking with DFG Collaborative Research Centers (CRC), Research Units (FOR), and Priority Programmes (SPP) in the Field of ESS (e.g. CRC754, CRC1294, TR32, TRR 172, TRR 181, FOR 2131, FOR 2589, SPP2115), where with increasing regularity data management and so called INF-projects are established. These data management projects are experienced in implementing processes and technologies over sub-disciplines and institutional borders. Hence, they provide good practice examples for the transition from project data to published data. The same holds for Cluster of Excellence in the field of ESS (EXC 2077, EXC 2037).

Additionally, the established relations on the level of data sharing to governmental institutions, will be reflected by the constitution of NFDI4Earth. Here, namely the Federal Institute for Geosciences and Natural Resources and the Federal Agency for Cartography and Geodesy already contributes to NFDI4Earth.

o internationally

Due to the large heterogeneity of communities within the international ESS there is a huge plethora of existing initiatives and infrastructures. Mapping these initiatives and building and strengthening links with the work of NFDI4Earth is an important task. Existing cooperations, e.g. with ESFRI initiatives such as ICOS or EPOS, but also with EOSC and many more offer potential here. The Global Earth Observation System of Systems (GEOSS) is of course in focus and also legal frameworks as the EU INSPIRE directive. Regarding geospatial data the Open Geospatial Consortium (OGC) is committed to making quality open standards for the global geospatial community. Many players in NFDI4Earth are engaged in OGC and the RDA (Research Data Alliance), others are members of the ICSU World Data System (WDS) or offer e.g. service for the International Association of Geodesy (IAG). NFDI4Earth is open for suggestions from the Belmont Forum.

o between the infrastructure facilities and the research community

There are already a number of initiatives and institutions that provide a strong link between research facilities and data infrastructures. One example is the DKRZ, which provides services for simulation-based climate and earth system research for researchers from different institutions. Another example is the Helmholtz Research Field Earth & Environment where a process was installed to improve the link between infrastructure and researchers.

o with respect to major networking topics

As already mentioned in the section "measures for user participation", the training of scientists of "all levels" is an essential building block for the success of NFDI4Earth. These trainings have EES-specific components, but can also benefit substantially from cooperation with networks in the field of training. The same applies to the field of standardization, software development and management. Here, too, NFDI4Earth aims to participate in new and existing networks.

· **Additional information**

For further developing the consortium NFDI4Earth a number of questions need to be addressed.

- NFDI4Earth includes already a large number of institutions and we are convinced that it will support perfectly for the ESS community. However, there are open questions on the role of our consortium within the planned structure of NFDI. How can different kinds and sizes of institutions be considered in the governance models for the NFDI at large? Is the request to have only one lead institution on the proposal helpful?
- We need more discussion and solutions on how to deal with cross-cutting activities. NFDI4Earth emphasizes subjects like education on data management and research software. It can be expected that more consortia will see this necessity. We see urgent need to show up pathways how generic aspects can be shared between the consortia without neglecting ESS specific needs.
- We see the need for a better advance definition of how interfaces between consortia can, even before the proposal is written.
- How can we offer services for other communities (e.g. geospatial information) and how can we express our needs on the services to other consortia?
- How can we organize the internal governance? Various models are currently under consideration, for the success of NFDI at large it will be crucial that consortia are led by comparable governances.
- NFDI4Earth strives to dynamically distribute an essential part of the funds made available via NFDI to additional institutional partners and activities during the course of the project and will have to establish transparent selection processes for this purpose. Guidance from the DFG is needed for this approach.