Letter of Intent - NFDI4SolidEarth

1 Non-binding letter of intent

This is a non-binding letter of intent as required as advance notification for a full proposal submission in 2020.

2 Formal details

2.1 Planned name of the consortium: NFDI for Solid Earth

2.2 Acronym of the planned consortium: NFDI4SolidEarth

2.3 Applicant

- Applicant institution
  Leibniz Institute for Applied Geophysics, Geozentrum Hannover, Stilleweg 2, Hannover
  Prof Dr. Manfred Frechen (deputy director).
- Spokesperson
  Dr. Thorsten Agemar, thorsten.agemar@leibniz-liag.de, Leibniz Institute for Applied Geophysics

3 Objectives, work programme and research environment

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

314 to 318

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

NFDI4SolidEarth addresses primarily geologists, geophysicists and other geoscientists. Geologists and geophysicists study the structure, composition and dynamics of the geological subsurface. Various invasive and non-invasive exploration methods are applied. The data obtained allow a differentiated examination of individual geological bodies and allow conclusions to be drawn about their structure and composition. In addition to answering fundamental questions on the history of the Earth's development, the focus is also on socially relevant topics such as sinkholes, earthquakes, drinking water, energy storage, geothermal energy, the search for a nuclear waste repository, resources extraction and the detection of landmines. As a rule, the data are processed in several process steps, so that many data sets can be available in different

versions depending on the degree of processing and the methods used. Detailed documentation of the data records is essential for reusability.

The development and use of information technologies has made great progress in the field of geosciences over the last 25 years. Today, measurement data is almost exclusively collected digitally and stored and processed in files or databases. In this context, an increasing degree of automation ("smart sensors") can be observed. Analog data (collections, drill records, magnetic tapes, etc.) are also recorded digitally in order to be preserved for research purposes. As a result of the often close international cooperation, more and more data is being made available to researchers all over the world via the Internet. For this reason, the geosciences are relatively far advanced in the implementation of the FAIR principles compared to other disciplines.

The growing availability of a multitude of samples, measurements, geologic interpretations and numeric simulations result in quickly growing amounts of data of increasing heterogeneity. A common characteristic of solid earth data is their spatial context.

These so-called georeferenced data do not only include maps and models, but every measurement or sample that can be assigned to a locality. As early as in the 1990s, the first geo-information systems were developed, which made it possible to select and retrieve data for research purposes from a map view. Databases also learned how to process geo-data and spatial queries. The technical means of displaying and retrieving spatial data via the Internet have developed rapidly over the last 25 years. The Open Geospatial Consortium (OGC, initially known as the Open GIS Consortium), founded in 1994, played a major role in this development.

Today, many existing information systems offer web-based GIS functionality that help users to find relevant data for their project area (e.g. GeotIS, FIS-Geophysik, map viewer of geological surveys etc.).

However, this development is by no means complete and there is still a strong need for the development of new user-friendly solutions for the provision and processing of solid earth data. Examples are reproducibility, quality control and scalability. At present, there are hardly any technical solutions and metadata standards that can be used to enable geoscientists to objectively understand the origin and quality of data. There are also no binding standards for documenting the quality of data and data repositories (however, this issue will become important if data publications are more strongly recognized as research achievements - analogous to the evaluation of scientific journals, for example on the basis of the citation index).

NFDI4SolidEarth thus addresses the growing complex digital requirements of the individual disciplines and sub-disciplines within solid earth sciences. NFDI4SolidEarth supports open science by promoting the use of OpenData and OpenSource software, and by fostering implementation of institutional data policies and data management plans that enable collaborative research and seamless data publication.

NFDI4SolidEarth will bring together scientists, research institutes, and geologic surveys to come up with new ideas and visions for better data management workflows, information systems, data formats and services. NFDI4SolidEarth stimulates innovation and cultural change in the ongoing digitalization of solid earth research. For the proposal, seven major task areas are envisaged:

1. ***Lead*** covers the overall coordination of the NFDI4Earth consortium, including the financial management. It will organize consortium meetings and set up a steering board. Further, it is responsible for the internal and external communication.

2. ***Data Services*** will develop innovative data services, tackle standardization issues for interoperability, develop algorithms for data visualization. One measure will be the standardization of interactive cross sections in web-GIS environments in collaboration with the OGC. In this case, the objective is the realization of interoperable cross section services across multiple servers with different coordinate systems. This task area aims at user-driven innovations for solid earth scientist resulting in new or extended OGC-services.

3. ***Data Formats*** covers all efforts in establishing, improving and promoting open data formats for solid earth data. For example, the BoreholeML markup language developed by the geological surveys in Germany is successfully used for the exchange of borehole data. However, in many situations, proprietary data formats are still de facto standards for solid earth data, especially in the 3D context.

4. ***Analytics*** will develop innovative data analytics frameworks and algorithms for data processing. Use cases could be for instance parallel simulations or the introduction of machine learning to seismic interpretation and attribute analysis.

5. ***Analogue to Digital*** does not only include the scanning of analogue recordings (e.g. seismic lines), lab reports, and borehole records before the paper disintegrates, it also includes the application of imaging methods to rock and mineral samples. The objective is to preserve data and knowledge about the subsurface for future research activities. This is a matter of due diligence we owe to future generations.

6. ***Quality*** covers all measures to assess or improve the usability, certainty, precision, and unambiguousness of data. Geoscientists often face the problem that complex data products like maps or 3D models do not come with a comprehensive documentation of all raw data included and processing steps applied. Sometimes it can also be relevant why a scientist has discarded data or processed it in some other way. Standardized workflows and documentation templates could help to assess the usability and optimization potential of complex data products. Further measures establish common criteria to evaluate the qualities of solid earth data. In the future, scientific data products will go through a peer-review process like today's scientific papers.

7. Metadata & Vocabularies focusses on improving existing metadata schemes (generic and specific to geophysics and geology) and facilitate metadata harvesting on existing information

systems and archives. Vocabularies (like for instance LithoLex of the state geological survey BGR[1]) are important tools for geoscientists who often face local or historic terms and definitions. Such vocabularies could be even more useful if connected to other information systems or databases. This task area will also tackle the consequent implementation of DOIs and ORCIDs in repositories. It has also strong links to INSPIRE, OGC-CWS, OWL (web ontology language), RDF (Resource Description Framework) and openScience initiatives like EOSC (European Open Science Cloud).

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 DFG web portal RIsources lists 37 research archives and research data repositories in the field of geosciences. However, only four of them relate to solid earth data:

- FIS-Geophysik (LIAG, Hannover)
- GeotIS (LIAG, Hannover)
- PANGAEA (AWI, Bremen)
- World Stress Map Project (GFZ, Potsdam)

There are other web portals for solid earth data, but many of them disappear after a few years in operation, such as Geomind. Nevertheless, NFDI4SolidEarth will continuously survey emerging databases and information systems for solid earth data with the objective to integrate them into NFDI.

The World Stress Map Project (WSMP) was initiated at the KIT in Karlsruhe and is now located at the GfZ in Potsdam. It offers validated data of the lithospheric stress field of tectonic plates derived from borehole observations. Its data are extremely relevant for the development of enhanced geothermal systems and the assessment of the slip tendency of faults. NFDI4SolidEarth intends to connect other information systems with WSMP and develop analytic frameworks on top.

PANGAEA is an archive of zip- or text files containing research data. It misses granular storage and retrieval options and a convenient map interface. Thus, it is the preferred archive for publishing project data presented in a research paper. The findability of PANGAEA data very much depends on the findability of the published paper. NFDI4SolidEarth will use PANGAEA to identify new relevant solid earth datasets for integration in advanced information systems for better visibility and re-usability.

[1] Bundesanstalt für Geowissenschaften und Rohstoffe

The web-based information systems FIS-Geophysics and GeotIS developed and operated by the LIAG are two advanced research data infrastructures with GIS features for solid earth data. The servers are hosted at the computing centre of the Geocentre of Hannover (BGR, LBEG[2] & LIAG). FIS-Geophysics is online since 2002 and holds geophysical subsurface data from industry, state authorities and research institutes. The coverage is primarily Germany. These data can be of great importance when evaluating sites for drinking water wells, geothermal plants or nuclear waste repositories, for example. Therefore, it is also an important data basis for the geothermal information system GeotIS. GeotIS provides public access to subsurface data relevant for deep geothermal energy utilization. The coverage is Germany and parts of Austria. It is online since 2007 and offers interactive visualization of geologic structures and geophysical data. GeotIS was the first web-GIS that offered interactive cross sections of the geologic subsurface.
The LIAG has a lot of experience in collaborative software development and its web-based information systems FIS-Geophysics and GeotIS are ideal candidates for pilot implementations and testing of enhanced data services, data formats, analytic frameworks and quality measures.

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

NFDI4SolidEarth has contacts to NFDI4Earth and NFDI4Objects. Hopefully some NFDI4SolidEarth co-applicants will also be participants of NFDI4Earth and vice versa. NFDI4Earth is complementary to NFDI4SolidEarth as NFDI4Earth focusses on atmosphere, hydrosphere and environment. Regarding the NFDI4Earth meetings in Hamburg, Potsdam and Göttingen, it has become clear that NFDI4Earth pursues predominantly generic approaches and is not open for solid earth disciplines. Thus, NFDI4Earth includes many disciplines but does not cover all earth system sciences. Despite the participation of the research institutions KIT, AWI, GFZ and Geomar as co-applicants within NFDI4Earth, the geophysicists and geologists are scarcely represented. This can already be seen from the fact that 108 relevant research data infrastructures were named in an NFDI4Earth internal survey, but not the World Stress Map Project or the Geo-Seas Project. NFDI4SolidEarth is taking action to close this gap with its own consortium and is ready to collaborate with NFDI4Earth.
NFDI4Objects and NFDI4SolidEarth have common interests, e.g. in the application and optimization of geophysical methods to detect hidden objects underground. The topic 3D-GIS can also be found in both consortia. NFDI4SolidEarth is confident that there will be a good exchange of ideas with NFDI4Objects regarding data services and visualization frameworks.

[2] Landesamt für Bergbau, Energie und Geologie

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.

The following cross-cutting topics are important from the NFDI4SolidEarth perspective:

- The self-organisation of consortia is a challenge. Since both top-down and bottom-up elements are always necessary for a functioning consortium, it is not easy to draft cooperation agreements. Thus, governance can be seen as an difficult issue.
- Progress in information technology and its applicability to enhanced research data infrastructures is also a relevant issue. Early identification of new developments and standards should be central to all consortia. This concerns metadata schemes, software/algorithms, interfaces, interoperability, coordinate systems, and data transfer.
- Data quality and uncertainty assessment especially in the 2D/3D context is an important issue.
- NFDI capacity building and education with activities towards research data literacy is also important.
- The promotion of OpenSource, OpenScience and OpenData is also recommendable.

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

NFDI4SolidEarth plans to contribute to all the above cross-cutting issues. Since NFDI4SolidEarth is still in the aggregation phase and will not submit an application until 2020 at the earliest, it is not yet possible to provide more detailed information here.