Interest Group for Geology & Geophysics I3G

The Interest Group Geology & Geophysics (I3G) within the framework of NFDI addresses the growing complex digital requirements of the individual disciplines and sub-disciplines within geological and geophysical sciences.

- I3G supports Open Science by promoting the use of Open Data and Open Source software, and by fostering implementation of institutional data policies and data management plans that enable collaborative research and seamless data publication.
- I3G brings together scientists, research institutes, and especially geologic surveys to develop new ideas and visions, and to establish workflows for better data management, interoperable information systems, data vocabularies, formats, standards and services.
- I3G stimulates innovation and cultural change in the ongoing digitalization of geological research.
- I3G will develop best practice examples of online information systems containing parameterized 3D geological models for spatial subsurface planning of economically and socially relevant geoscientific topics based on existing platforms.

The I3G focuses on the data infrastructure needs of geologists, geophysicists and other geoscientists. The data of interest covers the solid part of the Earth, the structure, composition and dynamics of the geologic subsurface. Various invasive and non-invasive exploration methods are included. The data obtained allow a differentiated examination of individual geological bodies and conclusions to be drawn about their structure, composition and properties. In addition to answering fundamental scientific questions on the history of the Earth's development, the focus is mainly on economically and socially relevant geoscientific topics such as georisks (e. g. sinkholes, earthquakes), drinking water, geoenergies, CO_2-storage, the search for a nuclear waste repository, and resources extraction.

However, this development is by no means complete and there is still a strong need for new user-friendly solutions for the provision, assessment and processing of geological and geophysical data. Due to the complexity of the subsurface and the large number of exploration methods and devices most geoscientists rely on specialized formats or software to work with data. This makes it difficult to use foreign data sets. Geoscientific research and subsurface spatial planning could be much more efficient and straightforward if these difficulties were overcome. Collaborations could be organized much easier if standardized web services and exchange formats tailored to the needs of geoscientists existed. Currently available solutions do not fully match geoscientific requirements. For instance, with some OGC-Services (e.g. WCS, WTS, WVS, 3DPS) it is possible to provide 3D data or at least views of 3D scenes as a web service to users, but
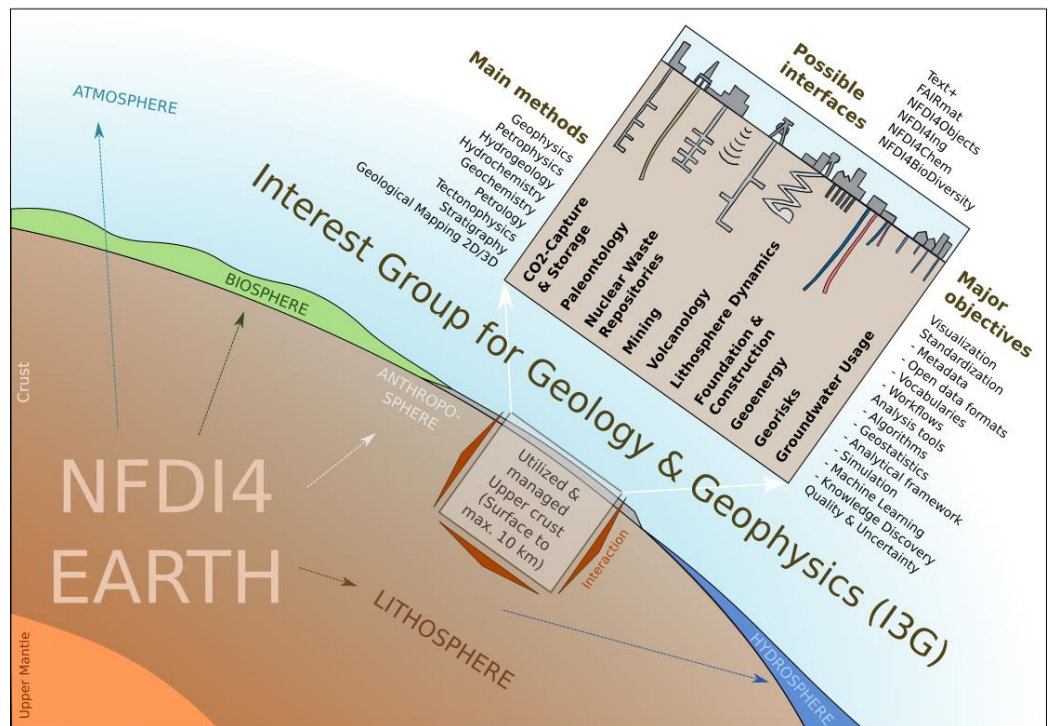

Figure 1: Main methods, topics and interfaces of I3G as part of NFDI4Earth.

Geoscientific background and state of the art

Industry, Geological Surveys, and research institutes have been collecting vast amounts of data (analogue and digital) over the past decades. Especially geophysical surveys generate large data sets. A common characteristic of all these data is their spatial context. Typically, geological and geophysical data relate to some subsurface geometry in a given coordinate system. However, geologists and geophysicists use various coordination systems and even the simple attribute 'depth' might be referenced as measured depth, true vertical depth, two-way-travel-time, or as stratigraphic horizon. Thus, spatial constraints used by geologists and geophysicists differ from those used by other scientists.

Today, many existing information systems offer web-based GIS functionality that help users to find relevant data for their project area. In recent years, some of these information systems learned to provide 3D contents. For example, it is possible to visualize 3D models of subsurface geology (sometimes including tunnels, mines, boreholes etc.) via WEB-GL in the browser and rotate them as desired. Or there is the possibility to draw a line on the map and thereby generate a geologic profile section of the subsurface.

the supported output data formats are not suitable for geological workflows. It would be very beneficial if geological 3D model data could be sliced and selected via a web service and integrated into 3D software and blended with data from other information systems in the future. In order to be able to interact with 3D geologic and geophysical data via the Internet as seamlessly as in the GIS world, a number of developments still need to take place. I3G has therefore set itself the target of joining all forces for a distributed research data infrastructure for 1D/2D/3D/4D data sets with a high level of interoperability, interactivity and scalability.

In the future, subsurface spatial planning will be increasingly dependent on being able to assess data properly. At present, few domains provide technical solutions and metadata standards that can be used to enable geoscientists to objectively comprehend the origin and quality of data or its applicability for a specific purpose. 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 a citation index. I3G will therefore support all measures for the introduction of community accepted concepts for data assessment and publication.

Algorithms and APIs being developed for geoscientific purposes could also be of great benefit to other disciplines since web based spatial data visualisation, assessment, validation and selection is a common exercise in Earth system science. Any code developed within NFDI should therefore be Open Source.

I3G focuses on the following topics:

1. Data Queries focusses on improving existing metadata schemes (generic and specific to geophysics and geology) and facilitate metadata harvesting on existing information systems and archives, as well as fuzzy search algorithms. This topic will tackle the consequent implementation of DOIs, ORCIDs and other national or international persistent identifiers in repositories. It has strong links to INSPIRE, OGC-CWS, OWL (web ontology language), RDF (Resource Description Framework) and Open Science initiatives like EOSC (European Open Science Cloud). However, a primary task will be the optimization of spatial data exploration techniques for faster and better query results. The major challenge within this task is the omnipresent time-depth-litho relationships in geoscience.

Figure 2: Overview of existing data and data infrastructure in comparison to an optimal workflow which should be developed and performed by the I3G

2. Data Services will tackle standardization issues for interoperability, develop algorithms for data visualization, selection, including the implementation of innovative concepts such as ontologies, artificial intelligence and machine learning. One option is to adjust OGC web services for 3D geologic data resulting in a new 3D web viewer and standardized interactive cross sections (horizontal and vertical) in web-GIS environments. This would enable geoscientists to blend data views in 2D/3D across multiple servers with different coordinate systems. This topic aims at user-driven innovations for geoscientist resulting in new or extended OGC-services.

3. Data Formats and Vocabularies covers all efforts in establishing, improving and promoting Open Data formats for geoscientific data. For example, the BoreholeML markup language developed by the Geological Surveys in Germany is successfully used for the exchange of borehole data. In the future, this XML-based data format is to be expanded so that geophysical measurements and hydraulic tests etc. can also be included. However, other exchange formats also exist (e. g. GeoSciML, SEP-3 etc.) and conversion tools need to be developed.

Other examples are vocabularies (e.g. LithoLex of the German Geological Survey BGR). These are important tools for geoscientists who are confronted with historic terms and cross-border deviating nomenclature. Connecting such vocabularies to information systems or databases would significantly increase their usefulness.

4. Analytics will develop innovative data analytics frameworks and algorithms for data curation and processing. Envisioned use cases are for instance geostatistic data evaluation, parallel simulations or the introduction of knowledge discovery and machine learning or artificial intelligence approaches for optimized or (semi-)automated seismic data interpretation and attribute analysis as well as other geoscientific multi-scale and -parameter datasets. Another field of interest is the development of artificial intelligence algorithms for the interpretation of logs and time series (e.g. seismology, petrophysics, sequence stratigraphy etc.).

5. Analogue to Digital focuses on the application of imaging methods to rock and mineral samples as well as outcrops and their inclusion in proper digital collections (e.g. IGSN, OutcropWizard). Machine learning algorithms could be applied to scanned logs and seismic recordings in order to decipher analogue data resources. The objective is to preserve data and knowledge about the subsurface for future research activities.

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 d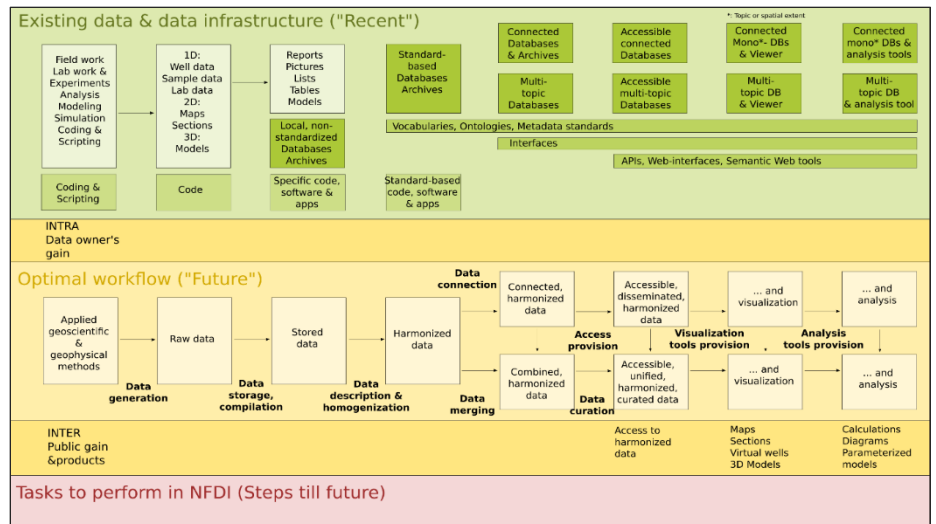o not come with a comprehensive documentation of all raw data included and processing steps applied and neither with a comprehensive quantified documentation of model uncertainty. It is relevant why a scientist has discarded data or processed it in some other way. Standardized workflows, versioning and documentation templates are envisaged to help to assess the usability and optimization potential of complex data products. Further measures establish common criteria to evaluate the qualities of geoscientific data. In the future, scientific data products will go through a peer-review process like today's scientific papers.

7. Multi-purpose 3D information systems will combine all of the above-mentioned tasks in the effort to develop a best practice example of an online information system for parameterized 3D geological models for spatial subsurface planning of economically and socially relevant geoscientific topics based on existing platforms (e.g. GeotIS, GeORG, WSM, etc.).

In Germany and abroad, a high quantity of research data on the geological subsurface can be accessed via web portals of governmental institutions. The new Geological Data Act ('Geologiedatengesetz') marks a turning point for the provision and use of geological and geophysical data for research purposes in Germany. This law will regulate the public provision of geological data by the State Geologic Services. A vast amount of industry data will become accessible for academic research. Since it is the responsibility of the geological surveys to request, collect and publish relevant measurement data of the geologic subsurface, a close collaboration between these federal authorities and research institutes is very important for mutual benefit and thus defined as target of I3G.

Members of I3G have demonstrated experience in software development (e. g. SeisViz3D, GeoReVi, seissol, GeotIS etc.). Most of the partners collaborate in joint projects, working groups or committees that relate to surveys, standards, methods and future concepts. The establishment of I3G does not require the establishment of new forms of exchange and cooperation. All I3G members have a clear idea of the tasks ahead and the nature of the collaboration. NFDI4Earth will be the chance to intensify this collaboration and provide additional benefits for geoscientific research and user communities.

Members of I3G will apply new solutions and concepts resulting from the topics described above to improve and interconnect their own public research data infrastructure, which are already widely used by scientists in Germany and abroad. I3G will continuously survey emerging databases and information systems for geoscientific data with the objective to integrate them into NFDI.

Contact:
Dr. Thorsten Agemar   thorsten.agemar@leibniz-liag.de
Dr. Kristian Bär   baer@geo.tu-darmstadt.de
Dr. Bianca Wagner   bwagner1@gwdg.de