

## **German marine seismic data access**

Christian Berndt<sup>1</sup>, Andreas Lehmann<sup>1</sup>, Christian Hübscher<sup>2</sup>, and Hanno Keil<sup>2</sup>

<sup>1</sup>GEOMAR Helmholtz Zentrum für Ozeanforschung Kiel

<sup>2</sup>Universität Hamburg

<sup>3</sup>Universität Bremen

24.7.2020

### **Abstract**

*Reflection seismic data are the paramount source of information for the subsurface structure as they provide the highest resolution of any geophysical technique. As such they have been used for a large variety of academic and commercial purposes. For many decades reflection seismic data were the largest data sets in Earth sciences, which created significant storage and archival problems and until today there is no international, European, or German data portal that is used throughout the community which hampers all new scientific projects that would like to use legacy data. Within this pilot project we will develop a unifying data infrastructure, kick off data archeology for existing data, and prepare for future archival of reflection seismic data from future research cruises. The initiative will serve primarily Germany's marine geophysics community as represented by AGMAR of the Fachkollegium Physik der Erde of DFG. The pilot for reflection seismic data contributes to the efforts of NFDI4Earth to establish a distributed infrastructure for data curation by harmonized data workflows with connections to international data repositories. Interfaces and storage solutions for data volumes in the order of GBs to TBs are urgently needed in many more scientific fields.*

### **I. Introduction**

Reflection seismic experiments are by far the most important source of information for answering a range of highly relevant societal and basic earth science questions, including:

- What is the role of the Earth system in past episodes of rapid global warming and can we trigger geological processes that accelerate global warming?
- How does the stability of submarine slopes evolve over time and what are the processes that lead to slope failure and associated tsunamis?
- How are natural resources distributed in the subsurface and which geological processes control their formation?
- How do crustal structures affect the slip mechanisms of mega-thrust faults that generate the largest earthquakes on Earth?

These are just few examples, but because of their high resolution compared to other geophysical methods they are basically used for all geoscientific work in which information about the subsurface is needed.

The data-challenge is two-fold. On one hand reflection seismic data have been acquired for the past 50 years without proper storage. The reason for this is that the data sets were traditionally among the largest data sets acquired in geophysics and storage capacity was limited. This means that still large amounts of data exist that are not stored in systematic ways and that will be lost to future generations of scientists. This is a disaster because it becomes increasingly difficult or impossible to collect these data again, because:

- it gets more difficult to obtain work permits for environmental reasons, marine mammal protection
- political changes prohibit access to some of the areas in which the data were acquired, e.g. Indonesia, Pakistan
- safety concerns prevent research vessels to enter specific sea areas, e.g. piracy in the Gulf of Arabia

It is not likely that this will change for the foreseeable future and precious data will be lost to mankind.

On the other hand, there is no central German repository for reflection seismic data. This means that all institutions that are collecting reflection seismic data cannot fulfill their obligation towards the DFG data stewardship recommendations in an optimal way but normally only the metadata are uploaded and in-house solutions are implemented for storing the raw data. This prevents efficient re-use of the data.

We envision a data portal for all existing and newly acquired German reflection seismic data that is accessible online and integrated into existing global initiatives and existing German solutions at other than academic stakeholders, e.g. the seismic database at the Bundesanstalt für Geowissenschaften und Rohstoffe (BGR).

This would allow the geophysical community to access reflection seismic data efficiently and provide a boost in particular to global and overview studies. This would lead to a step increase in scientific output and enable completely new approaches such as big data analysis, e.g. to make prediction of petrophysical properties worldwide.

## ***II. Pilot description***

Seismic data pose a particular problem compared to other earth science data: the raw data cannot be used directly but the data have to be processed by experienced processors first which contains an element of subjectivity and access to very expensive processing software that is constantly being developed further. This means that different types of data, i.e. meta data, raw data and various processing steps as well as their chosen parameters have to be stored and made accessible. Only a complete documentation of the processing steps will ensure reusability and reproducibility. At the same time the data volumes are still comparably large, i.e. 10s of GB for one survey and one processing step. This requires a sensible choice of data types to be stored and access to large-scale infrastructure.

During the pilot phase we will first evaluate the different initiatives in other countries, e.g. at UTIG, Austin or LDEO, Woods Hole, SeaDatanet, and at BGR in order not to duplicate efforts and develop a metadata standard to ascertain operability. BGR has committed to support this pilot study with their expertise garnered from their in-house data management program which organized their reflection seismic data from 2005 to 2015.

Once this has been established, we will make recommendations regarding the infrastructural implementation, e.g. a connection to PANGAEA data publisher or GFZ seismological data archive with additional interfaces or separate solutions.

In the second phase of the pilot study we will start to compile a list of all existing reflection seismic data sets that have to be archived. This will take place in close collaborations with the geophysical institutes that are organized in the Arbeitsgemeinschaft Marine Geophysik (AGMAR) of the Fachkollegium Physik des Erdsystems (FKPE) of DFG. AGMAR has already provided a letter of support to NFDI4Earth to support this initiative.

The solution will be the establishment of a data portal that can host all German reflection seismic data that have been collected over the past 50 years and those that will be collected on future expeditions. This will vastly improve the current situation that risks a partial loss of the existing data and hampers scientific progress.

### ***III. Relevance for the NFDI4Earth***

Expected users of the seismic data portal are the national and international scientific community – primarily in geology and geophysics but also in adjacent field such as climate research. These communities will benefit from better access to valuable data, especially when there is a standard for metadata description for e.g. geographic exploration and well described data formats. Other sub-branches in the Earth System Sciences can potentially benefit from the workflows established to allow a standardized data flow from expedition to archive especially for large data sets.

A portal solution for reflection seismic data addresses all aspects of FAIR data: data are **findable** at one place, described with a known and standardized set of metadata. The data are **accessible** online via the portal. **Interoperability** is achieved by setting up a service to access data from different sources and all this leads to possible **reuse** of data that have been hidden for several decades.

This pilot addresses both the very beginning of the data lifecycle, e.g. acquisition of metadata (via DSHIP) and data during a cruise and the dataflow to store raw data in the repository, and the 'end' of the lifecycle: by using persistent identifiers for these data sets, the reference of data in scientific papers allows easy reuse of the data for further analyses.

The pilot for reflection seismic data contributes to the efforts of NFDI4Earth to establish a distributed infrastructure for data curation by harmonized data workflows with a connection to international repositories. Interfaces and storage solutions for data volumes of the order of GBs to TBs are urgently needed in many more scientific fields such as climate simulations or image

processing. This pilot can be used as blueprint for them. The pilot will foster further links throughout the marine geophysics community and we envisage to feed this initiative into NFDI4Earth geophysics interest group.

## IV. Deliverables

The first deliverable is a report which provides an overview of the world-wide initiatives for reflection seismic data storage, as well as the metadata, and data products that have to be archived and made available in the new portal using persistent identifiers. Specific emphasis will be on the adaptability of the BGR scheme for the whole community and which changes have to be implemented in order to adapt it in NFDI4Earth.

The second deliverable is a road map document “German reflection seismic data portal” including a prototype implementation. The road map will set out a realistic plan on how to build up the portal, conduct the data archiving, and automatically integrate it with the DSHIP acquisition system to provide seamless integration for future research cruises. The road map document will detail the implementation and maintenance strategy by the marine and geophysical community within the NFDI4Earth framework including time line, partners and required funding.

## V. Work Plan & Requested funding

	1	2	3	4	5	6	7	8	9	10	11	12
International review	x	x	x									
Portal strategy			x	x	x	x						
Adaptation of BGR database					x	x	x					
Specification of data / meta data requirements					x	x						
Compilation of existing data						x	x	x	x			
Development of roadmap									x	x	x	
Full-scale proposal for funding											x	x
Community meetings	x			x			x			x		

We are applying for twelve months E13 salary to carry out the pilot work plus 2000 EUR for organizing the community workshops and visit each of the involved institutions (Uni Kiel, AWI, Uni Bremen, Uni Hamburg, BGR, IOW).