International Research: **World Databases**

Databases as fundamental tools for interdisciplinary research: PANGAEA Network for Geological and Environmental Data and the World Data Center for Marine Environmental Sciences (WDC-MARE)

H.-J. Wallrabe-Adams[1], M. Diepenbroek[1], H. Grobe[2], and R. Sieger[2]

[1] *MARUM - Centre for Marine Environmental Sciences , Bremen, Germany*
[2] *Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany*

Introduction

Increasing amounts of analytical data in geosciences and an increasing cooperation of various scientific disciplines makes it necessary to establish efficient data information systems. Such a system should be able to handle all kinds of geoscience data.

InterRidge maintains several databases containing ridge related data especially metainformation on vessels and vehicles, research cruises, location and general information on known hydrothermal vent areas, vent biological information, ridge related references and the linked Petrological Database of the Ocean Floor (PETDB) at Lamont-Doherty Earth Observatory. Like PETDB WDC-MARE/PANGAEA contains measured and calculated numerical data, textual data and pictures. Among others, parameter groups like geochemistry of rocks, sediments, water (*e.g.* CTD), compositional data of sediments like grain size, particle association, fossil distributions, physical properties etc. are available. All data sets are georeferenced (latitude, longitude, depth/elevation/age) and include information about the source (publication), investigator(s), research area, and parameters inclusive pertaining analytical/calculation methods.

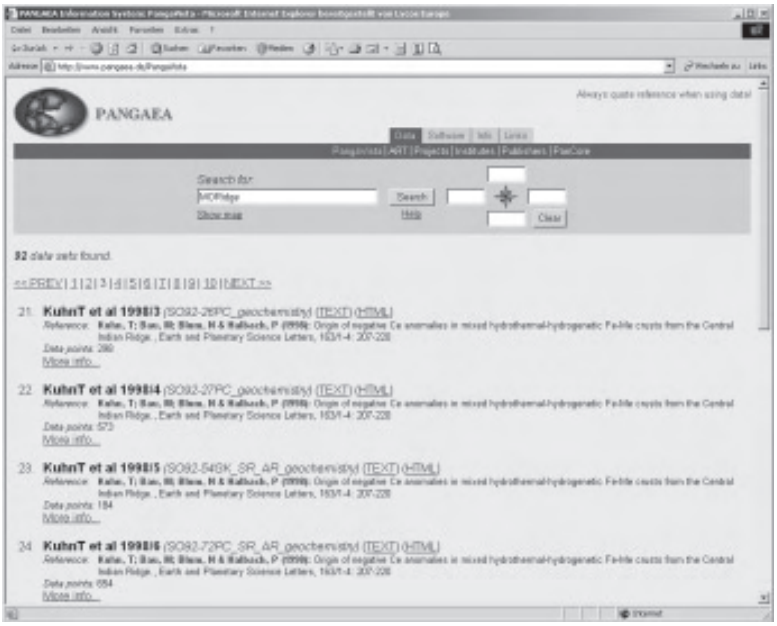

Figure 1. PangaVista is a simple search engine, which allows retrieving desired data sets by seeking matches to user-specified keywords and geographical constraints. Here a number of data sets were found for the query 'MORidge'. The map shows the corresponding data points

WDC-MARE/PANGAEA data information system

The World Data Center for Marine Environmental Sciences (WDC-MARE) was founded in 2000 and is a member of the ISCU World Data Center system [1]. As an operating platform for WDC-MARE PANGAEA is used, which is an information system for processing, long term storage, and publication of georeferenced data related to earth sciences. Essential services supplied by WDC-MARE / PANGAEA are project data management and the distribution of visualization and analysing software. Organization of data management includes quality check and publication of data and the dissemination of metadata according to international standards.

The challenge of managing the heterogeneous and dynamic data of environmental and geosciences was met in the PANGAEA system through a flexible data model which reflects the information processing steps in the earth science fields and can handle any related analytical data (Diepenbroek *et al*., 1999; in press). The basic technical structure corresponds to a three tiered client/server architecture with a number of clients and middleware components controlling the information flow and quality. On the server side a relational database management system (RDBMS) is used for information storage. The web-based clients include a simple search engine (PangaVista, Fig. 1) and a sophisticated data mining tool (ART). The client used for maintenance of information contents is

optimised for data management purposes. Analysing and visualization of metainformation and analytical data is supported by a number of software tools.

With its comprehensive graphical user interfaces and the built in functionality for import, export, and maintenance of information, PANGAEA is a highly efficient system for scientific data management and data publication.

The PANGAEA and the WDC-MARE are operating on a long term basis. The institutional frame is supplied by MARUM in cooperation with the Alfred Wegener Institute (AWI), Bremerhaven. Several people are responsible for the technical and scientific organization and development. The data management totals about six fulltime scientists. Data management services on an international level are available since 1996. Until the beginning of 2002 WDC-MARE / PANGAEA was a partner in 34 projects covering all fields of environmental sciences (*http://www.pangaea.de/Projects/*).

The WDC-MARA/PANGAEA tasks in detail

1) "Pooling" of project relevant data and metainformation (data description), which includes the acquisition and incorporation of data into the information system as well as a harmonisation of data and a final quality control. Part of this task will also be "data mining" for and revision of already existing data relevant to ridge research.
2) General access for individual scientists or any interested groups. The pooled data will be accessible through the general web interfaces of PANGAEA and a project specific web page, which reflects the actual status of the data management at any time, will be available. The functionality of PANGAEA also allows for the distribution of restricted data, which is important during the runtime of the project. Project members can define flexible access rights to their data sets.
3) Long term archiving and publication of data. Publication of data includes linkage of data with corresponding publications and attachment of data specific abstracts/comments. For further propagation metadata will be disseminated into international clearinghouses as *e.g.* the Global Change Master Directory (GCMD).
4) Support for processing and synthesis of data, in particular the harmonization of project data and preparation of compiled data sets for visualization and analysis in GIS.

The InterRidge Next Decade-workshop emphasized the importance and fundamental meaning of databases for global and interdisciplinary research efforts like InterRidge. Scientists of different research fields need an effective information exchange system to combine data from different kinds of ridge related studies (*e.g.* to combine data from hydrothermal vent biota with fluid chemistry). Extensive experiences in project data management , *e.g.* the opportunity to manage JGOFS final data synthesis as a most recent example, encourage the WDC-MARA/PANGAEA group to offer InterRidge the data handling and publication of cruise and post-cruise data. This is thought as a tool for InterRidge researchers to overview existing data, plan future research, and distribute the own results for more efficient scientific communication. The database is easily accessible and usable by common internet browsers. WDC-MARE/PANGAEA database is available at *http://www.wdc-mare.org*, *http://www.pangaea.de* or via *http://www.pangaea.de/Projects/INTERRIDGE/* . This link is accessible from the InterRidge website through the "IR Databases" menu.

WDC-MARA/PANGAEA will be the major data management system for the future German Ridge activities (Priority program of the Deutsche Forschungsgemeinschaft DFG).

Last but not least, the helpfulness of a database depends on the readiness of scientist to publish and archive their results in the database. We encourage everybody to contact the WDC-MARA/PANGAEA group (*info@pangaea.de*) to discuss data storage and data exchange.

References

[1] Panel on World Data Centers. Guide to the World Data Center System, issued by the Secretariat of the ICSU Panel on World Data Centers, *http://www.ngdc.noaa.gov/wdc*, 1996.

Diepenbroek, M., H. Grobe, M. Reinke, R. Schlitzer, and R. Sieger. Data management of proxy parameters with PANGAEA. In: Fischer, G., and Wefer, G. (eds.), Use of Proxies in Paleoceanography - Examples from the South Atlantic, Springer, Berlin, Heidelberg, 715-727, 1999.

Diepenbroek, M., H. Grobe, M. Reinke, U. Schindler, R. Schlitzer, R. Sieger, and G. Wefer, (in press). PANGAEA – an Information System for Environmental Sciences. Computers and Geosciences, *http://www.pangaea.de/Paper/CG/*

The deadline for *InterRidge News* 12(1) is:

17 March, 2003

Format specifications can be found at:
http://www.intridge.org/irn.htm