Archiving data from earth system research - a show case

Hannes Grobe & Rainer Sieger
What is PANGAEA?

Pangaea is an information system for georeferenced data from basic research on the earth system. Data are stored in a relational database which is accessible on the Internet. The system is open to individual scientists as well as projects to preserve data and to make it public available.

Intention and operation of Pangaea is comparable to a library - a public electronic library for research data.

The system faces any technical challenges, which a modern information system on the Internet should have. It is continuously developed and adopted to new technical and scientific requirements and standards.
Digital „Library“
in
Open Access
for
Georeferenced Data
from
Earth & Environment
Does science need a *library* for data?
MOTIVATION

- Open Access to scientific results
- Safeguarding ‘Good Scientific Practice’
- Added value through integration
- Overview on ‘what exists’
- Persistent identification and availability
- Reporting to funding organization
- Personal record + credit
- BackUp

Libraries preserve the knowledge of mankind!
Good scientific practice in research and scholarship
European Science Foundation (ESF), 2000

Data accumulation, handling, and storage

36. Data are produced at all stages in experimental research and in scholarship. Data sets are an important resource, which enable later verification of scientific interpretations and conclusions. They may also be the starting point for further studies. It is vital, therefore, that all primary and secondary data are stored in a secure and accessible form.

37. Institutions may pay particular attention to documenting and archiving original research and scholarship data. Several codes of good practice recommend a minimum period of 10 years, longer in the case of especially significant or sensitive data. National or regional discipline-based archives should be considered where there are practical or other problems in storing data at the institution where the research was conducted.
Initiatives & Protocols

Open Access

Budapest Open Access Initiative

The Budapest Open Access Initiative arises from a small but lively meeting convened in Budapest by the Open Society Institute (OSI) on December 1-2, 2001. The purpose of the meeting was to accelerate progress in the international effort to make research articles in all academic fields freely available on the

Conference on
Open Access to Knowledge in the Sciences and Humanities
26 - 22 Oct 2003, Berlin

Berlin Declaration

Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities
OECD Principles and Guidelines for Access to Research Data from Public Funding
Who are the hosts of PANGAEA?

Pangaea is operated by the
(1) Alfred Wegener Institute for Polar and Marine Research,
member of the Helmholtz Association of National Research Centres,
funded by the Federal Ministry of Education and Research
and the
(2) Center for Marine Environmental Sciences at the University of Bremen
with support of the Department of Geoscience and the
research center ocean margins,
funded by the German Research Foundation.

Both institutions have committed to long-term operate Pangaea and the
World Data Center for Marine Environmental Sciences.
OPERATING INSTITUTIONS

Center for Marine Environmental Sciences, Bremen

Alfred Wegener Institute for Polar and Marine Research
Bremerhaven
WORLD DATA CENTER SYSTEM OF ICSU

ICSU
International Council for Science

WDC

PANGAEA - Publishing Network for Geoscientific & Environmental Data
Final data report for projects

CD/DVD with data and local search engine

Description and further information in a booklet

Distribution through 300 libraries with focus on marine research

WDC-MARE Reports
Integrated Data Sets of the DFG Research Project SFB 313
Environmental Change: The Northern North Atlantic
(Veränderungen der Umwelt: Der nördliche Nordatlantik)

Hannes Grobe, Michael Diepenbrock, Priska Schäfer, Jörn Thiede & Gerold Wefer

WORLD DATA CENTER FOR MARINE ENVIRONMENTAL SCIENCES
Alfred Wegener Institute for Polar and Marine Research, Bremerhaven
MARUM Center for Marine Environmental Sciences, Bremen

http://www.wdc-mare.org
How are metadata and data organized?
# Data Classification

<table>
<thead>
<tr>
<th>Technical</th>
<th>Scientific Level</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>numbers</td>
<td>primary (raw)</td>
<td>counts</td>
</tr>
<tr>
<td>text</td>
<td>secondary</td>
<td>percentage</td>
</tr>
<tr>
<td>objects</td>
<td>tertiary</td>
<td>calculations interpretations</td>
</tr>
</tbody>
</table>
GEO-CODE & META-DATA

**when?**
- date/time
- age

**what?**
- parameter
- [unit]

**how?**
- method

**where?**
- latitude
- longitude
- ice, water, air, sediment, object...

**who?**
- investigator
- reference

123.456
- text

PANGAEA - Publishing Network for Geoscientific & Environmental Data
## Georeferenced Data

### Data Table

<table>
<thead>
<tr>
<th>Value</th>
<th>Latitude</th>
<th>Longitude</th>
<th>3. Spatial dim</th>
<th>Date/Time</th>
<th>Age (kyr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.555</td>
<td>-67.6421</td>
<td>18.7659</td>
<td>0.3</td>
<td>2004-03-04</td>
<td></td>
</tr>
<tr>
<td>23.7</td>
<td>-67.6421</td>
<td>18.7659</td>
<td></td>
<td></td>
<td>5.7</td>
</tr>
<tr>
<td>123.456</td>
<td>-67.6421</td>
<td>18.7659</td>
<td>112.45</td>
<td></td>
<td>1230</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Binary Object

<table>
<thead>
<tr>
<th>Latitude</th>
<th>Longitude</th>
<th>3. Spatial dim</th>
<th>Date/Time</th>
<th>Age (kyr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-67.6421</td>
<td>18.7659</td>
<td>0.3</td>
<td>2004-03-04</td>
<td></td>
</tr>
</tbody>
</table>
Client-server system
with three-tiered architecture

1. Import (Curator/Editor)
   - Import Server

2. Database Server

3. Web Service/Server
   - Export (User)

Internet
How can I find and download data?
Data Access

- **Pangaea** search engine
- **ART** (Advanced Retrieval Tool) > curators only
- **DDI** (Direct Download Interface) > dynamic query
- **PanCore** > Metadata search
- **DOI** (Digital Object Identifier) > persistent link
- **Web service** > exchange with portals (ISO, DC, XML)
- **Data-Warehouse** > retrieval & compilation
I. no need
II. file collection
III. working data base
IV. data center or repository
V. portals and grid
PANGAEA® – standard interfaces for metadata

- Data management & long-term archiving
- Catalogues
- Protocols
- Catalogues
- Frontends / portals

- marshaller
- RDB
- PANGAEA
  - Index
  - WFS (SOAP/WSDL)
  - OGC catalogue service
  - ISO19xxx
  - Dublin Core
  - DIF
  - Darwin Core
  - STD-DOI

- XSLT
- ISO19xxx
- Dublin Core
- DIF
- Darwin Core
- DOI registration

- WS (SOAP/WSDL)
- OAI-PMH
- DiGIR
- ISO690
- DOI registry

- GeoPortal. Bund®
- IODP
- D-GRID
- CARBOOCEAN
- EUR-OCEANS
- GCMD
- OAIster
- Google
- HGF Fedora
- OBIS
- GBIF
- TIB National Library
Dissemination of data and metadata via search engines, library catalogs & portals

Search Engines
- Google
- GFF
- planktonnet

Projects
- OBIS
- GBIF
- planktonnet

Meta Systems
- GEOPortal.BUND

Portals
- TIB
- IODP
- CATster

Library Catalogs
- ePIC
How can I use data from Pangaea in my own application?
DATA MINING

Pan2Applic

Mean Sea Level Pressure [Pa]
DATA WAREHOUSE

< Globigerina bulloides >
Ocean Data View (ODV)
What type of data can I archive/find in PANGAEA?
## MAJOR PROJECTS

<table>
<thead>
<tr>
<th>International</th>
<th>EU</th>
<th>National</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSRN</td>
<td>OMARC</td>
<td>rcom</td>
</tr>
<tr>
<td>JGOFS</td>
<td>CarboOcean</td>
<td>SIRRO</td>
</tr>
<tr>
<td>WOCE</td>
<td>EurOceans</td>
<td>HISTRA</td>
</tr>
<tr>
<td>EPICA</td>
<td>HERMES</td>
<td>ARCOD</td>
</tr>
<tr>
<td>IODP</td>
<td>EPOCA</td>
<td>DFG/BMBF</td>
</tr>
</tbody>
</table>
Examples from Geoscientific Research

- Sediment profile
- Seismic profile
- Mineral distribution
- Geological map
Examples from Environmental Research

- Images
- Distributed samples
- Oceanographic profiles
- Times Series
Air photos

doi:10.1594/PANGAEA.323540
Sea-bed photos

doi:10.1594/PANGAEA.319877
International Polar Year  (1882-1883)
NO DATA FOUND?

might not be the problem of the database...
Contact

info@pangaea.de

This presentation is available at hdl:10013/epic.31845.d001