Making Data Accessible: Suggestions from the Scientific Community

Gerold Wefer, Universität Bremen
21. September 2007
The Ocean in the Earth System
Sea-bed photos

doi:10.1594/PANGAEA.319877
Sediment core documentation

doi:10.1594/PANGAEA.108079
NW – SE seismic multichannel seismic line over Chapopote/ Campeche Bight

MCS Line M672a-GeoB06-116 - 10 m CDP spacing, FD migrated
Data types in PANGAEA

- Profiles -> [doi:10.1594/pangaea.103958](https://doi.org/10.1594/pangaea.103958)
- Time series -> [doi:10.1594/pangaea.32348](https://doi.org/10.1594/pangaea.32348)
- Sea bed photos -> [doi:10.1594/PANGAEA.31987](https://doi.org/10.1594/PANGAEA.31987)
- Distributes samples -> [doi:10.1594/pangaea.51749](https://doi.org/10.1594/pangaea.51749)
- Complex data -> [doi:10.1594/PANGAEA.108079](https://doi.org/10.1594/PANGAEA.108079)
- Air photos -> [doi:10.1594/PANGAEA.323540](https://doi.org/10.1594/PANGAEA.323540)
- Audio record -> [doi:10.1594/PANGAEA.339110](https://doi.org/10.1594/PANGAEA.339110)
PANGAEA –
our data information system
Data policy for using the information system PANGAEA as Open Access archive, data library and publishing system

World Data Center for Marine Environmental Sciences (WDC-MARE)
Alfred Wegener Institute for Polar and Marine Research (AWI), Bremerhaven & Center for Marine Environmental Sciences (MARUM), Bremen, Germany

The aim of this policy is to facilitate operation and use of the information system PANGAEA - Publishing Network for Geoscientific & Environmental Data by the research community. This policy recognises the benefits of providing free and open access to good quality data from earth and environmental sciences for future use in global change studies, research projects, and operational services such as portals and search engines. The operating institutes encourage the widest possible use of the Pangaea library, in order to best realise its potential value.

Principles

- The guiding principle of the PANGAEA - Publishing Network for Geoscientific & Environmental Data is free and open access to its content by research and education communities in non-commercial activities. This is in line with data policies of the IOC, the WDC System and the OECD.

doi:10.1594/PANGAEA.327791
Challenges

- Long-term operation
  - hardware, software, backup, network

- Simple and flexible data model
  - extensible and adjustable to evolving science

- Scientific and technical standards
  - for science and web integration

- Scalable access
  - depending on different user needs

- User driven and controlled
  - to avoid a technical end in itself
PANGAEA / WDC-MARE - organisation

AWI
Computer Center

MARUM
Univ. Bremen

technical & scientific organization

Project & data managers

Bremen network internet

www.wdc-mare.org www.pangaea.de
Operating Institutions

Center for Marine Environmental Sciences, Bremen

Alfred Wegener Institute for Polar and Marine Research
Bremerhaven
PANGAEA® – dissemination of data and metadata via portal networks
PANGAEA - services & activities

- Data management
- Data publication
- Data infrastructures (networking)
  - Data portals, networking
  - Citable data sets
  - Distribution through 300 libraries
    - Accompanied by CD/DVD with data and local search engine
  - Accompanied by CD/DVD with data and local search engine
  - Accompanied by CD/DVD with data and local search engine

WDC-MARE Reports 0003 2005

International Collection of JGOFS (Joint Global Ocean Flux Study)

Rainer Sieger, Hannes Grobe, Michael Dispenbroek, Uwe Schindler, Reiner Schlitzer (Editors), JGOFS DMTT & IPO

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

www.wdc-mare.org  www.pangaea.de
PANGAEA - Statistics (7/2007)

Total number of data sets  
Data items

541,468
> 1.8 billions

www.pangaea.de
WDC-MARE - World Data Center for Marine Environmental Sciences
Biogeochemistry, Circulation, and Life of Present and Past Oceans

http://www.wdc-mare.org/

PANGAEA - Publishing Network for Geoscientific and Environmental Data

http://www.pangaea.de/
Word Data Center for Marine Environmental Sciences
Biogeochemistry, Circulation, and Life of Present and Past Oceans

Objectives

The World Data Center for Marine Environmental Sciences (WDC-MARE) is aimed at collecting, scrutinizing, and disseminating data related to Global Change in the fields of environmental oceanography, marine geosciences, and marine biology. It focuses on georeferenced data (numeric, text, and any kind of binary objects) using the PANGAEA information system as its long-term archiv and publication unit.

WDC-MARE is maintained by the Alfred Wegener Institute for Polar and Marine Research (AWI) and the Center for Marine Environmental Sciences (MARUM). Additional support is provided by the Research Center Ocean Margins (rcom).

Latest News [Archive]

2005-12-31
Data archiving of DFG project 'Ocean Gateways' completed.

2005-11-01
The first three issues of the WDC-MARE Reports are printed and distributed to about 300 libraries worldwide.

2005-09-13
The data library PANGAEA used by WDC-MARE for archiving provides 250,000 data sets consisting of about half a billion data points.

http://www.wdc-mare.org
World Data Center System of ICSU

established during the International Geophysical Year 1957/58
Principles of the World Data Center System

- operation for the benefit of the scientific community
- long-term archival
- freely available
- data exchange
Project STD-DOI “Data Publishing”

Technische Informationsbibliothek Coordination
(TIB, Hannover)

World Data Center Climate - WDC-Climate
(MPI, Hamburg)

World Data Center for Marine Environmental Sciences - WDC-MARE
(AWI/MARUM, Bremen)

World Data Center for Remote Sensing - WDC-RSAT
(DLR, Oberpfaffenhofen)

GeoForschungszentrum - WDC-TERRA?
(GFZ, Potsdam)

http://www.std-doi.de
MARUM – WDC MARE is developing SEDIS Phase I for IODP-MI: central metadata catalogue listing of all IODP, ODP, and DSDP datasets
Summary

• Acceptance of a data system stands or falls with simplicity
• Data must be accompanied by standard descriptions
• Data must include in its description a usable citation
• Data storage must be managed by established centers
Long term preservation

Lifetime of storage media (years):

- Hard disk: 5
- CD / DVD: 20
- Tape: 30
- Paper: > 100
- Papyrus: > 1000

Stone of Rosette
Recommendations

• We have already well established data information systems. What we need are more data submitted to the centers.

• Funding agencies should formulate their data policies with appropriate explanations and regulations and special funding should be provided.