An introduction to the Data Library PANGAEA®

Stefanie Schumacher, Rainer Sieger & Hannes Grobe (2010)
Data sharing and archiving

Nature:
Vol 461, 10 September 2009

doi:10.1038/461145a
Empfehlungen der Kommission "Selbstkontrolle in der Wissenschaft"

Vorschläge zur Sicherung guter wissenschaftlicher Praxis
Januar 1998

Empfehlung 7

Primärdaten als Grundlagen für Veröffentlichungen sollen auf haltbaren und gesicherten Trägern in der Institution, wo sie entstanden sind, für zehn Jahre aufbewahrt werden.
Pangaea is an Open Access data library for earth system research. Data are stored georeferenced in space and time in a relational database and a tape archive.

The data content is accessible on the Internet via a search engine, a data warehouse and web services.

The system is open to any scientist or project to archive and publish data.
History & Milestones

- 1987  Core repository database
- 1989  SEDI/SEDAT proprietary predecessor
- 1994  SEDAN/SEPAN relational predecessor
- 1996  PANGAEA
- 1998  www.pangaea.de  
  each dataset can be identified, shared, published and cited by using a Digital Object Identifier (DOI)
- 2001  WDC-MARE
- 2004  OAI and DOI
- 2006  Data citation, portal software
- 2008  Data warehouse
- 2009  Elsevier-Partnership
PANGAEA—an information system for environmental sciences

Michael Diepenbroek, a, b, Hannes Grobe, a, Manfred Reinke, a, b, Uwe Schindler, c, Reiner Schlitzer, b, Rainer Sieger, b, and Gerold Wefer, a

a Center for Marine Environmental Sciences (MARUM), University Bremen, Bremen 28334, Germany
b Alfred Wegener Institute for Polar and Marine Research, Bremerhaven 27515, Germany
c Physics Department, University of Erlangen-Nuremberg, Erlangen 91058, Germany

Received 23 March 2001; revised 20 April 2001; accepted 5 May 2001. Available online 20 September 2002.

Abstract

PANGAEA is an information system for processing, long-term storage, and publication of georeferenced data related to earth science fields.
DOI – Digital Object Identifier

Is a character string used to uniquely identify an electronic document or object.

The DOI for a document is permanent, whereas its location and other metadata may change.

Is resolved by a doi-resolver: http://dx.doi.org/

Example:

doi:10.1594/PANGAEA.737668

http://dx.doi.org/10.1594/PANGAEA.737668
Who are the hosts of PANGAEA?

(1) Alfred Wegener Institute for Polar and Marine Research (AWI)

member of the Helmholtz Association of National Research Centres
funded by the Federal Ministry of Education and Research (BMBF)

&

(2) Center for Marine Environmental Sciences (MARUM)

at Bremen University
funded by the German Research Foundation (DFG)

Both institutions have committed to long-term operate PANGAEA and the World Data Center for Marine Environmental Sciences (WDC-MARE)
Publication of data with PANGAEA
Ontogenetic effects on stable carbon and oxygen isotopes in tests of live (Rose Bengal stained) benthic foraminifera from the Pakistan continental margin

Stefanie Schumacher, Frans J. Jorissen, Andreas Mackensen, Andrew J. Gooday, and Olivier Pays

Laboratory of Recent and Fossil Bio-Indicators (BIAF), Angers University, 2 Bd Lavoisier, 49045 Angers Cedex 01, France
Laboratory of Marine Bio-Indicators (LEBIM), Ile d’Yeu, Ker Chaton, France
Alfred Wegener Institute for Polar and Marine Research, Am Allen Hafen 26, 27568 Bremerhaven, Germany
National Oceanography Centre, Southampton, European Way, Southampton SO14 3ZH, United Kingdom
LEESA, Ecology and Conservation Biology group, Angers University, 2 Bd Lavoisier, 49045 Angers Cedex 01, France

Received 11 December 2008; revised 10 June 2010; accepted 17 June 2010. Available online 25 June 2010.

Abstract
Compilation of ozone sonde profiles from the Antarctic
Georg-Forster-Station from 1985 to 1992

G. König-Langlo and H. Gernandt
Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany
Abstract. On 22 May 1985 the first balloon-borne ozone sonde was successfully launched by the staff of Georg-Forster-Station (70°46’S, 11°41’E). The subsequent weekly ozone soundings mark the beginning of a continuous investigation of the vertical ozone distribution in the southern hemisphere by Germany.

The measurements began the year the ozone hole was discovered. They significantly contribute to other measurements made prior to and following 1985 at other stations. The regular ozone soundings from 1985 until 1992 are a valuable reference data set since the chemical ozone loss became a significant feature in the southern polar stratosphere.

The balloon-borne soundings were performed at the upper air sounding facility of the neighboring station Novolazarevskaya, just 2 km from Georg-Forster-Station. Until 1992, ozone soundings were taken without interruption. Thereafter, the ozone sounding program was moved to Neumayer-Station (70°39’S, 8°15’W) 750 km further west.

Data coverage and parameter measured

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Short Name</th>
<th>Unit</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Altitude</td>
<td>Altitude</td>
<td>m</td>
<td>height above mean sea level</td>
</tr>
<tr>
<td>Date/Time</td>
<td>Date/Time</td>
<td></td>
<td>universal time code (UTC)</td>
</tr>
<tr>
<td>Longitude</td>
<td>Longitude</td>
<td></td>
<td>at launching point</td>
</tr>
<tr>
<td>Latitude</td>
<td>Latitude</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geone, partial pressure</td>
<td>Press</td>
<td>hPa</td>
<td></td>
</tr>
<tr>
<td>Temperature, air</td>
<td>Temp</td>
<td>°C</td>
<td></td>
</tr>
<tr>
<td>Wind direction</td>
<td>Wind</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wind speed</td>
<td></td>
<td>m/sec</td>
<td></td>
</tr>
</tbody>
</table>

Correspondence to: G. König-Langlo

(ger.koenig-langlo@awi.de)

Published by Copernicus Publications.
Final data report for projects

CD/DVD with data and local search engine

Description and further information in a booklet

Distribution through 270 libraries with focus on marine research
What type of data are archived in PANGAEA?
# Major Projects

<table>
<thead>
<tr>
<th>International</th>
<th>EU</th>
<th>National</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radiation</td>
<td>Pollen</td>
<td>Marine environment</td>
</tr>
<tr>
<td>JGOFS</td>
<td>CarboOcean</td>
<td>Tree rings</td>
</tr>
<tr>
<td>Oceanography</td>
<td>Ocean acidification</td>
<td>Data archaeology</td>
</tr>
<tr>
<td>Ice cores</td>
<td>HERMES/Hermione</td>
<td>DFG/BMBF</td>
</tr>
<tr>
<td>Marine geology</td>
<td>EPOCA</td>
<td></td>
</tr>
</tbody>
</table>

[http://www.pangaea.de/projects/]
Examples from Geoscientific Research

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

- Images
- Distributed samples
- Hydrographic profiles
- Times Series
Examples from Antarctic Research

- Southern Ocean Atlas
- Ozone profiles
- Sediments and Rocks
- CRP Cape Roberts Project
- Archive of Underwater Imaging
- EPICA European Project for Ice Coring in Antarctica
JGOFS
Joint Global Ocean Flux Studies

---

1. **Mackey, DJ (2003):** ADCP current measurements at cruise FR8/93 (southbound)
   - *Reference: CSIRO (2000):* Australian Equatorial JGOFS data set,

---

Please quote reference and citation when using data!
Sea-bed photos

   References: Ragné-Gill, JM, Gutt, J, Gutt, J, Clarke, A et al. (2004). Antarctic shallow-water mega-epibenthos: shaped by circumpolar dispersion or local conditions I. Marine Biology Gutt, J, Arnitz, WE, Balgurias, E et al. (2003): Diverse approaches to questions of diversity: German contributions to benthos studies around South America and Antarctica, Gayana Gutt, J, Pienzberg, D (2003): Scale-dependent impacts of catastrophic disturbances by grounding icebergs on the diversity of Antarctic benthos, Marine Ecology Progress Series
   Size: unknown
doi:10.1594/PANGAEA.319877 - Score: 60% - Similar datasets

   Size: unknown
doi:10.1594/PANGAEA.319885 - Score: 60% - Similar datasets

   References: Gutt, J, Arnitz, WE, Balgurias, E et al. (2003): Diverse approaches to questions of diversity: German contributions to benthos studies around South America and Antarctica, Gayana Gutt, J, Pienzberg, D (2003): Scale-dependent impacts of catastrophic disturbances by grounding icebergs on the diversity of Antarctic benthos, Marine Ecology Progress Series Gutt, J, Starmans, A (2001): Quantification of iceberg impact and benthic recolonisation patterns in the Weddell Sea (Antarctica), Polar Biology
   Size: unknown
doi:10.1594/PANGAEA.319884 - Score: 60% - Similar datasets

   doi:10.1594/PANGAEA.319877
Sediment core documentation

doi:10.1594/PANGAEA.108079
Geological map

doi:10.1594/PANGAEA.138789
Bathymetry

doi:10.1594/PANGAEA.351142
Meteorological observations

doi:10.1594/PANGAEA.269619
Data  Archeology
The PANGAEA web server operates the Mirror Site for the Ocean Drilling Program (ODP) in Europe.

http://odp.pangaea.de

Röhl et al. 2000
doi:10.1594/PANGAEA.57539
Geo-code & meta-data

**when ?**
date/time or age

**what ?**
parameter [unit]

**how ?**
method

**where ?**
latitude, longitude

**text**
123.4

**ice, water, air, sediment, object...**

**who ?**
investigator

**reference**
... no data without metadata
no metadata without data ...
Empty archives

Most researchers agree that open access to data is the scientific ideal, so what is stopping it happening? Bryn Nelson investigates why many researchers choose not to share.

In 2003, the University of Rochester in New York launched a digital archive designed to preserve and share dissertations, preprints, working papers, photographs, music scores—just about any kind of digital data the university's investigators could produce. Six months of research and marketing had convinced the university that a publicly accessible online archive would be well received. At the time of the launch, the university librarians were worried that a flood of uploaded data might swamp the available storage space.

Six years later, the US$200,000 repository lies mostly empty. Physicists, mathematicians and computer scientists use arXiv.org, operated by Cornell University in Ithaca, New York; the International Council for Science's World Data System holds data for fields such as geophysics and biodiversity; and molecular biologists use the Protein Data Bank, GenBank and dozens of other sites. The astronomy community has the International Virtual Observatory Alliance; geoscientists and environmental researchers have Germany's Publishing Network for Geoscientific & Environmental Data (PANGAEA),

"We got the software up and running and said 'Give us your stuff'. That's
Workflow in data publishing

- Provision of data (PI)
- Import to PANGAEA (curator)
- Proof-Read (PI)
- Corrections (curator/editor)
- Peer review (reviewer ?)
- Publication with DOI & citation
Keep in mind:

Submit your data to PANGAEA before your manuscript is in press

Reference in the paper to your data by doi:

For supplementary data see doi:10.1594/PANGAEA.472241

Data can be pass-word protected until the paper is published

Data formats: Preferred format for data tables is TAB-delimited TEXT-files (ASCII), submitted as ZIP-archive, or excel-format

Curator for GLOMAR-related data is Lydia Gerullis

See also: http://wiki.pangaea.de/wiki/Main_Page
Link to your data

for Elsevier publications a link on their web page
Final take-home message

use Digital Object Identifier (DOI) or Handles (hdl) instead of URLs

PANGAEA datasets can be identified, shared, published and cited by using a Digital Object Identifier (DOI)

DOI-resolver: http://dx.doi.org

This presentation is available at hdl:10013/epic.36320.d001
How can I find and download data?

www.pangaea.de

Examples