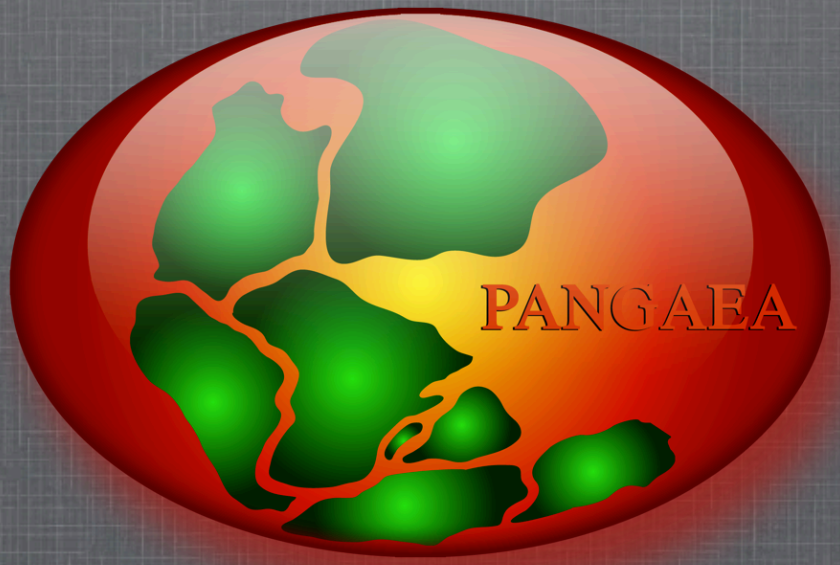
Archiving data from earth system research - a show case



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 continously developed and adopted to new technical and scientific requirements and standards.

WWW.PANGAEA.DE

Digital „Library“

in

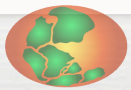
Open Access

for

Georeferenced Data

from

Earth & Environment

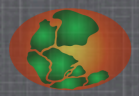


Does science need a *library* for data ?

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

OECD Principles and Guidelines for Access to Research Data from Public Funding

OECD

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

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

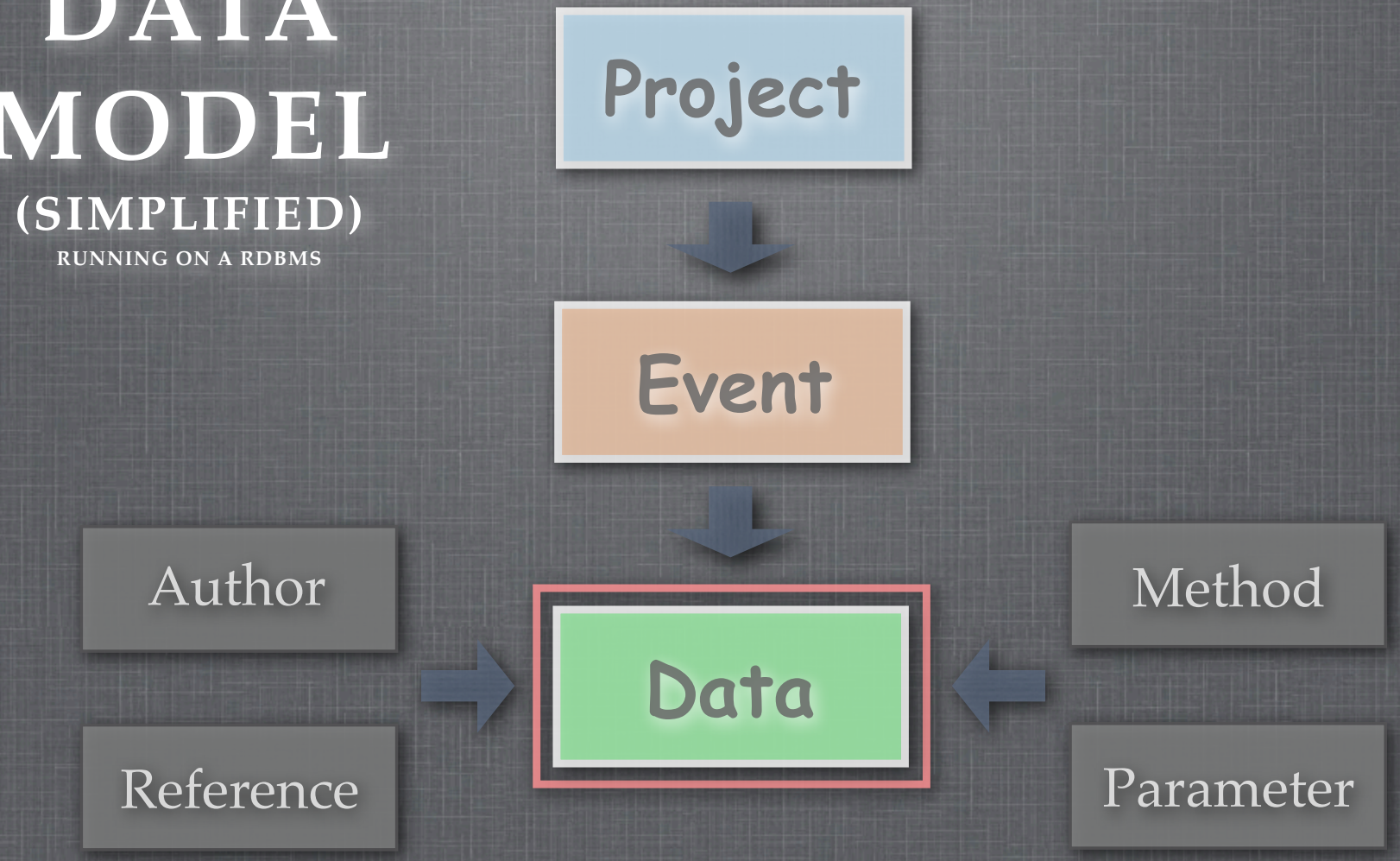
http://www.wdc-mare.org

How are metadata and data organized ?

DATA MODEL

(SIMPLIFIED)

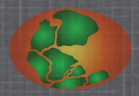
RUNNING ON A RDBMS



DATA CLASSIFICATION

technical
numbers
text
objects

scientific level	*example*
primary (raw)	*counts*
secondary	*percentage*
tertiary	*calculations* *interpretations*

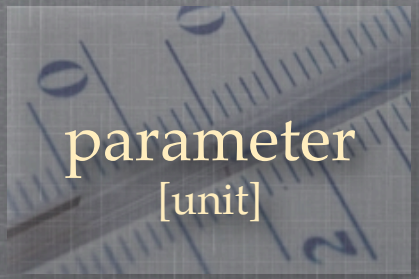


GEO-CODE & META-DATA

when ?

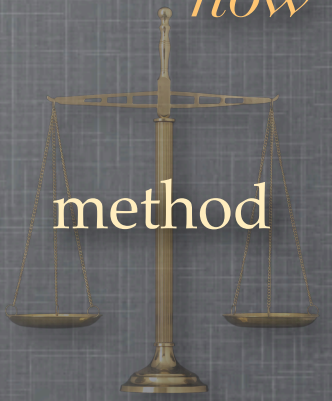
date / time
age

what ?



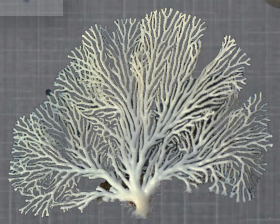
parameter
[unit]

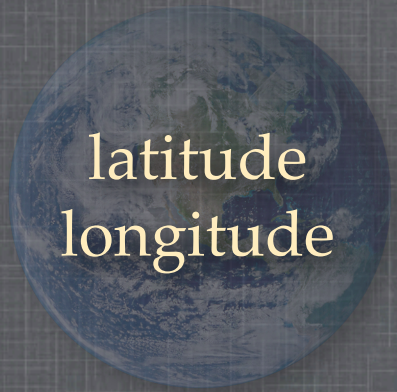
how ?



method

123.456 text



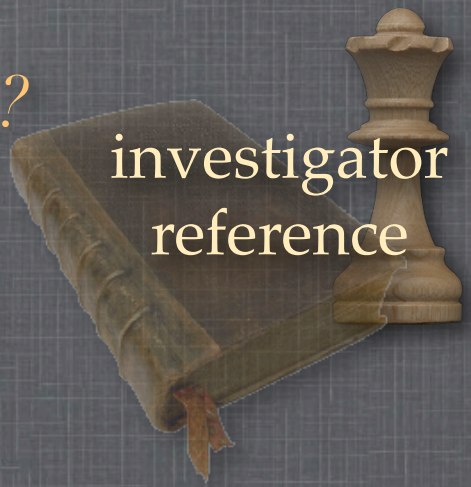


latitude
longitude

where ?

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

who ?



investigator
reference

GEOREFERENCED DATA

Data table

Value	Latitude	Longitude	3.Spatial dim	Date/Time	Age (kyr)
0.555	-67.6421	18.7659	0.3	2004-03-04	
23.7	-67.6421	18.7659			5.7
123.456	-67.6421	18.7659	112.45		1230
...	...				

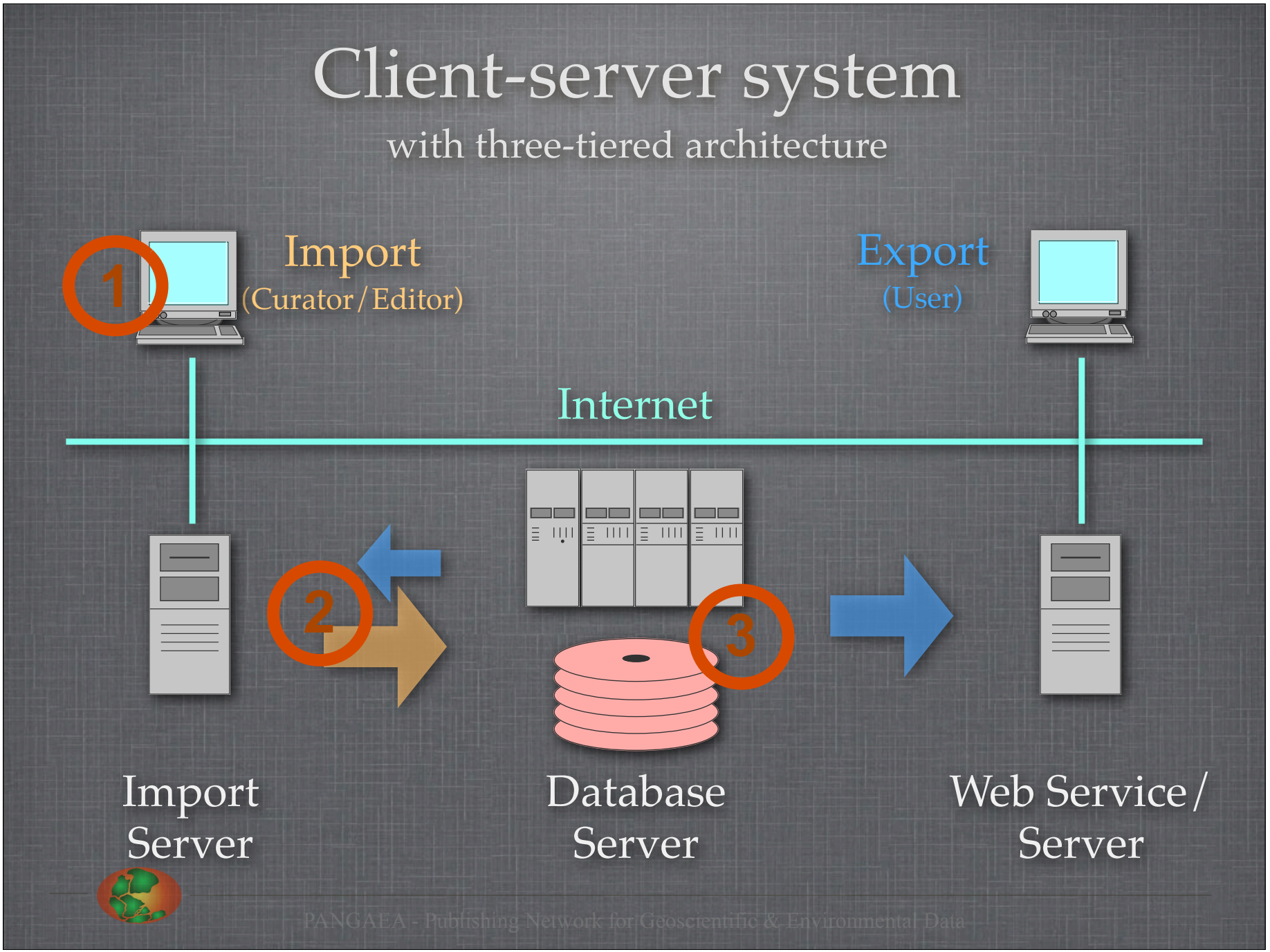
Binary object

123.456
text
file
...

Latitude	Longitude	3.Spatial dim	Date/Time	Age (kyr)
-67.6421	18.7659	0.3	2004-03-04	

Client-server system

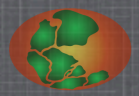
with three-tiered architecture



How can I find and download data ?

Data Access

- **Pangaea** search engine
- **ART** (Advanced Retrieval Tool) > curators only
- **DDI** (Direct Download Interface) > dynamic querie
- **PanCore** > Metadata search
- **DOI** (Digital Object Identifier) > persistent link
- **Web service** > exchange with portals (ISO, DC, XML)
- **Data-Warehouse** > retrieval & compilation



Search engine

Search for:

Search

Show map

Help

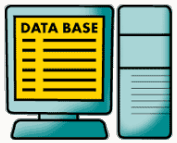
Info Software Advanced search

ENTROPIE IN DATA ARCHIVING

I. no need

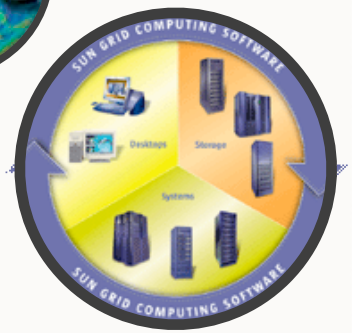
II. file collection

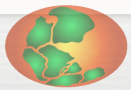
III. working data base



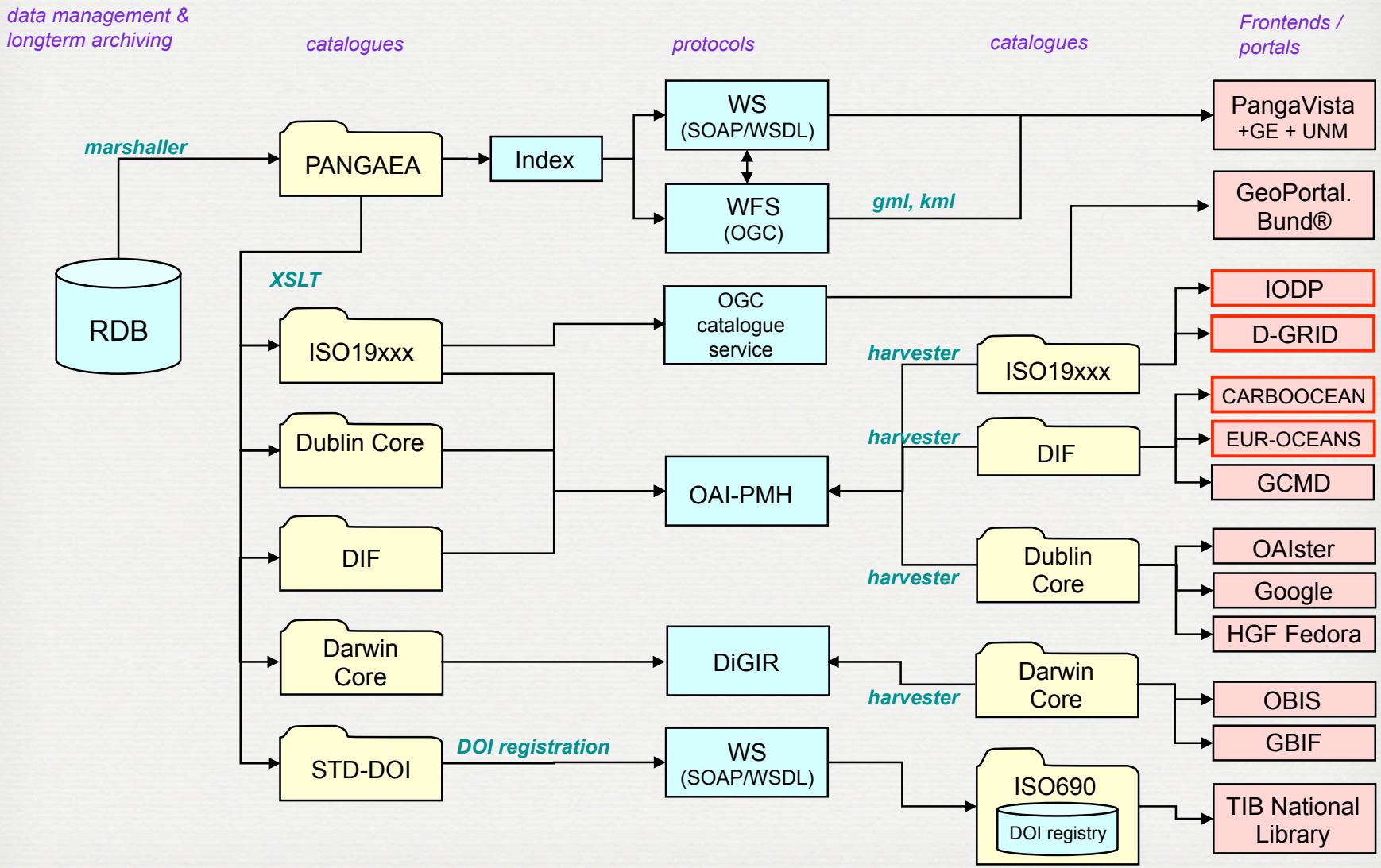
IV. data center or repository

V. portals and grid

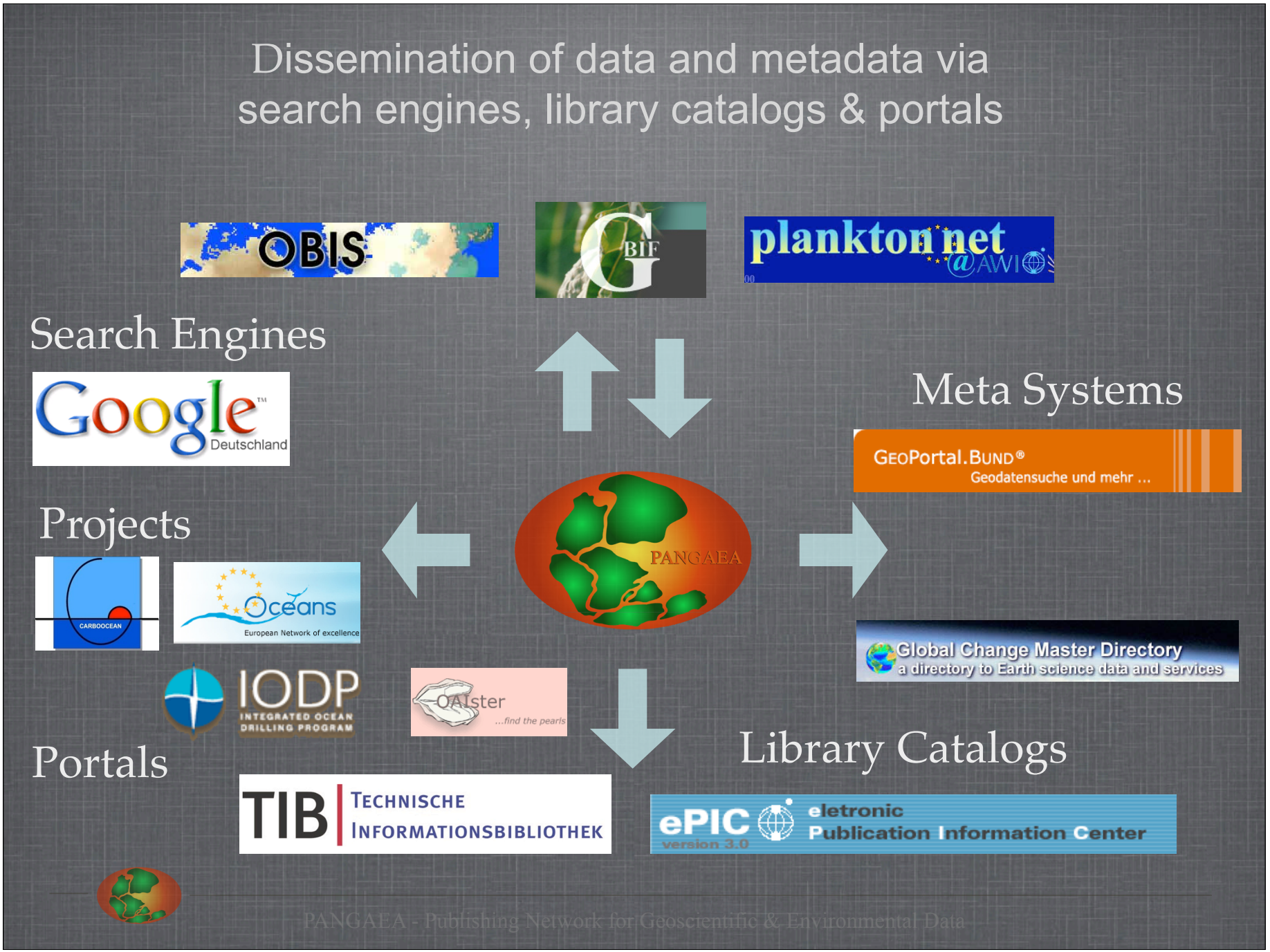




PANGAEA® – *standard interfaces for metadata*



Dissemination of data and metadata via search engines, library catalogs & portals



How can I use data from Pangaea

in my own application ?

DATA MINING

Parameters

- ☐ Ammonium [µmol/l] *(param52)*
- ☐ Nitrate [µmol/l] *(param756)*
- ☐ Phosphate [µmol/l] *(param758)*
- ☐ Oxygen [µmol/l] *(param754)*
- ☐ Silicate [µmol/l] *(param755)*

Geographic Coverage

Clear

Temporal Coverage

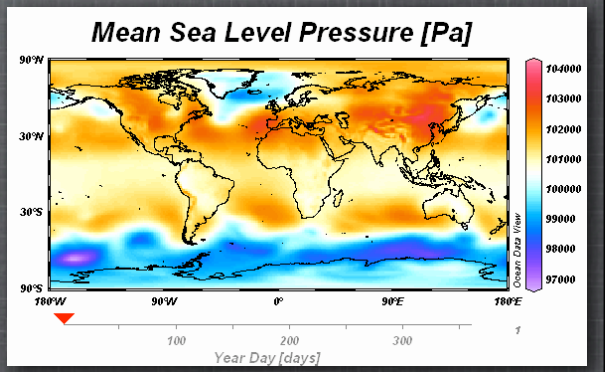
Start date:
End date:
Clear

Water Depth

Minimum [m]:
Maximum [m]:
Clear

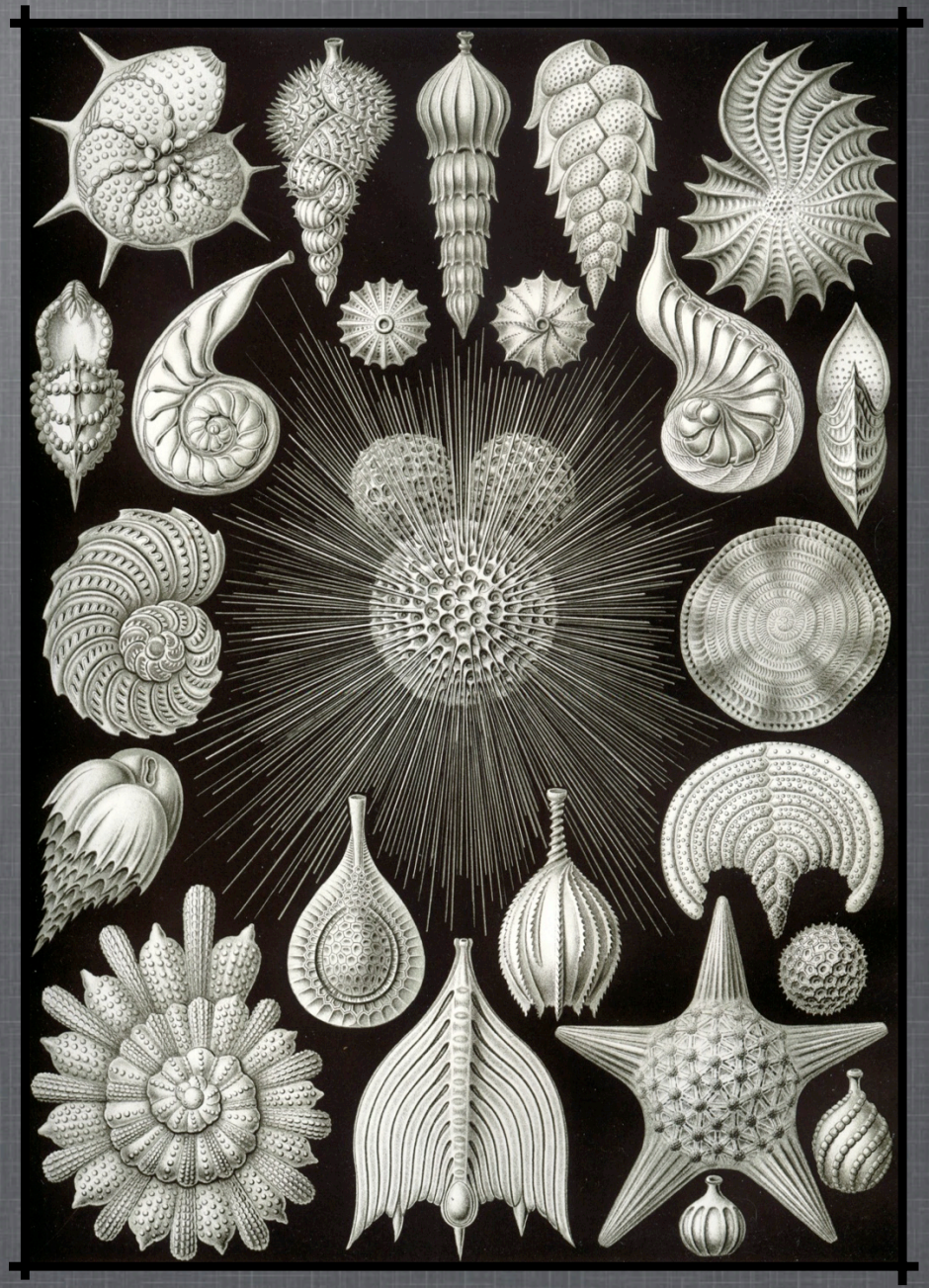
Start IQ Query

Pan2Applic

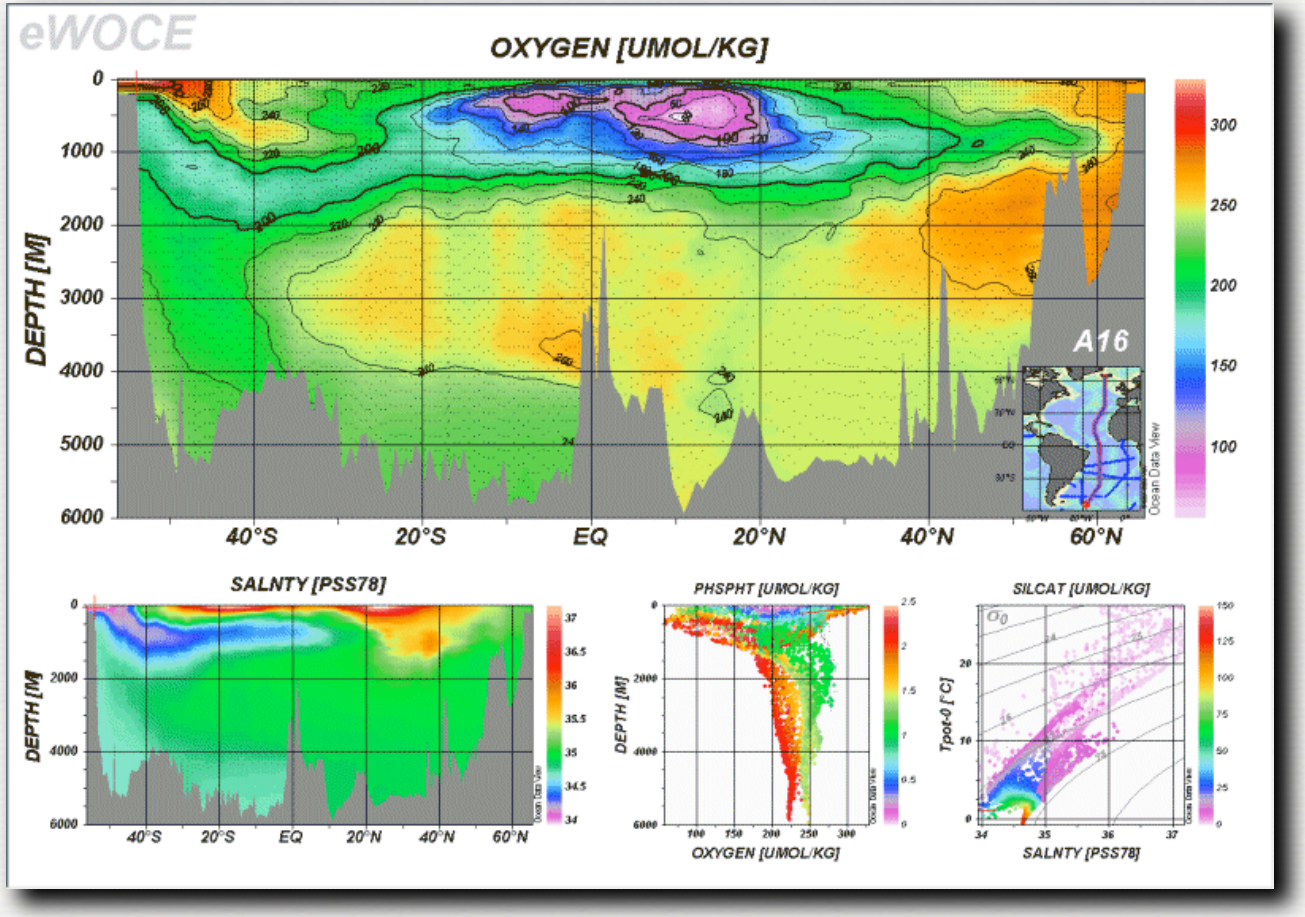


DATA WAREHOUSE

< Globigerina bulloides >



Ocean Data View (ODV)

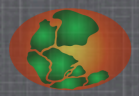


What type of data can I archive/find in PANGAEA ?

MAJOR PROJECTS

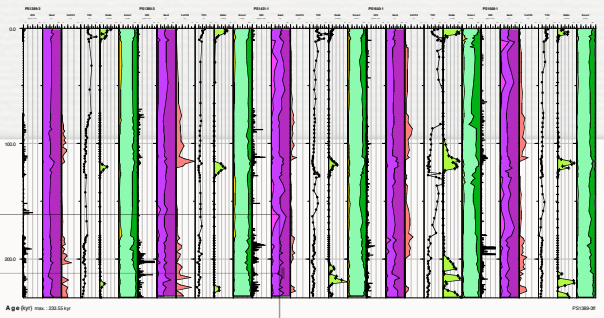
International	EU	National
BSRN	OMARC	rcom
JGOFS	CarboOcean	SIRRO
WOCE	EurOceans	HISTRA
EPICA	HERMES	ARCOD
IODP	EPOCA	DFG/BMBF

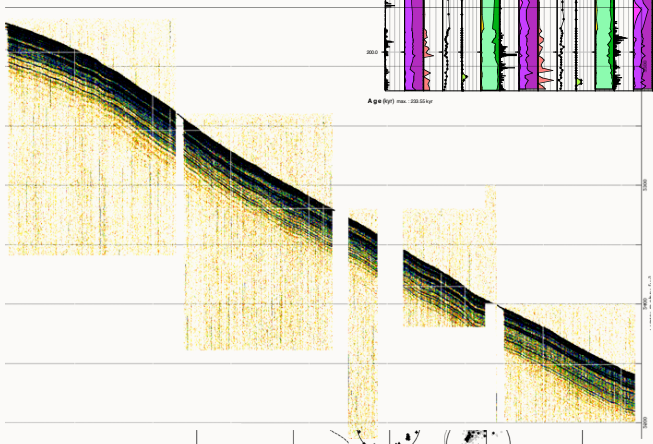
http://www.pangaea.de/projects/

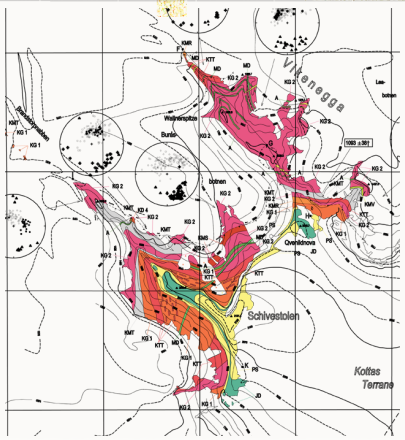


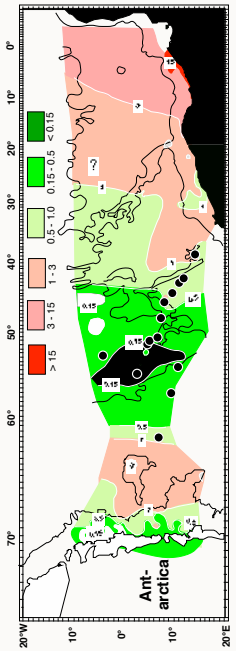
Examples from Geoscientific Research

- Sediment profile
- Seismic profile
- Mineral distribution
- Geological map



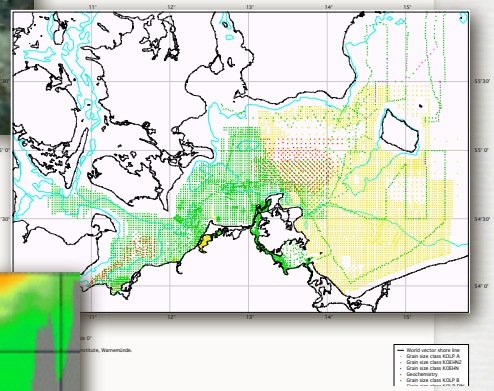


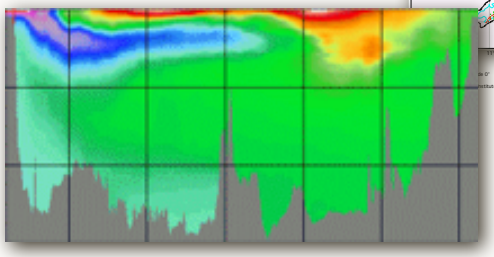


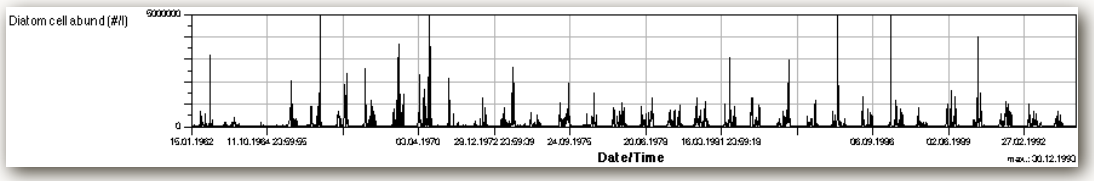


Examples from Environmental Research

- Images
- Distributed samples
- Oceanographic profiles
- Times Series



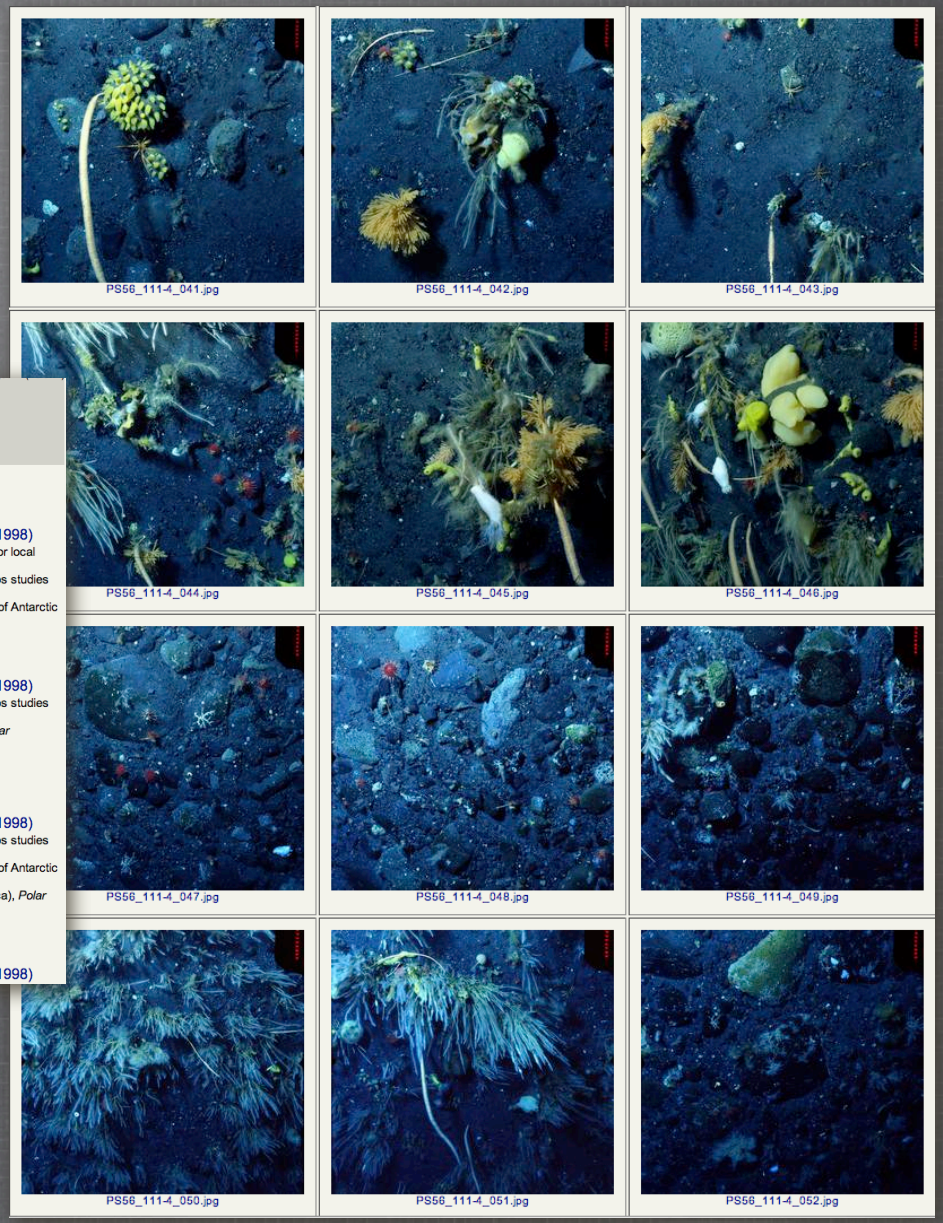




Air photos

doi:10.1594/PANGAEA.323540

Sea-bed photos



Search for:
sea-bed
Search
Show map
Help
Clear

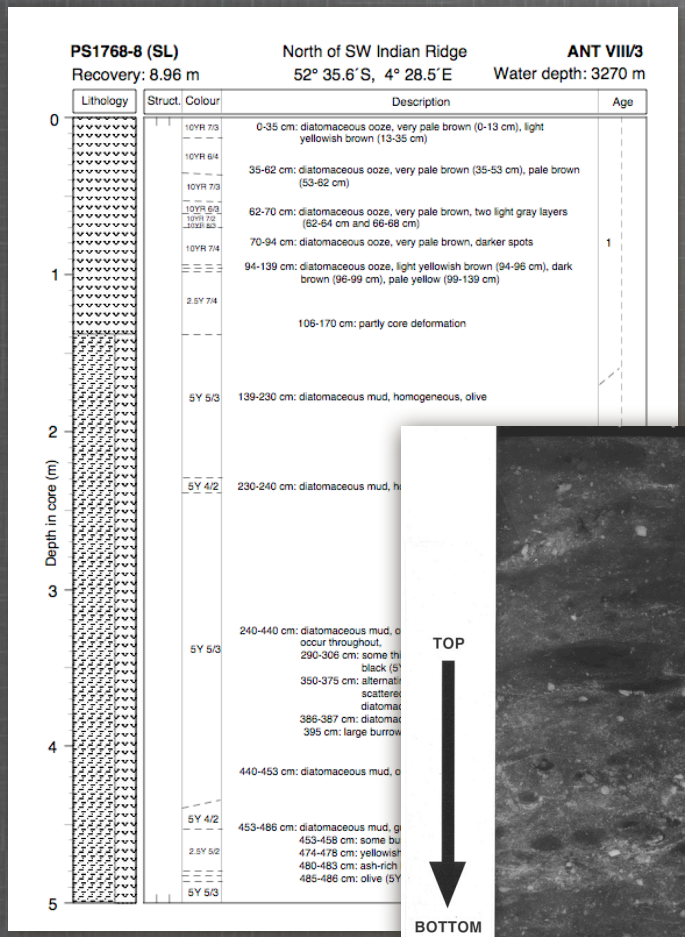
189 datasets found!

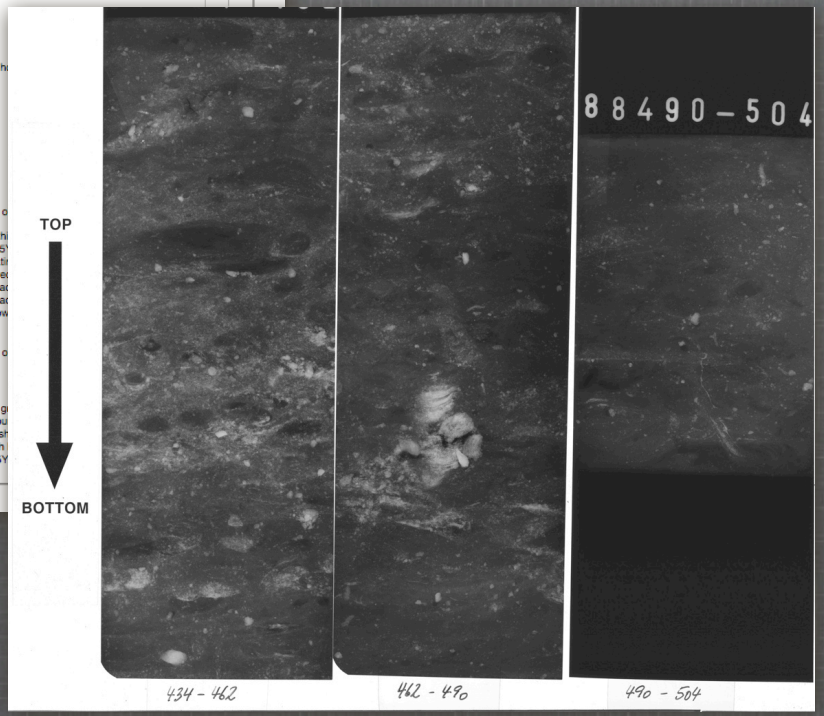
<< PREV | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | **10** | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | NEXT >>

1. **Gutt, J (2004):** Sea-bed photographs (benthos) from the Weddell Sea along ROV profile PS48/281 (©AWI, Gutt 1998)
   Reference: **Raguá-Gil, JM; Gutt, J; Clarke, A et al. (2004):** Antarctic shallow-water mega-epibenthos: shaped by circumpolar dispersion or local conditions?, *Marine Biology*
   Gutt, J; Arntz, WE; Balguerias, E et al. (2003): Diverse approaches to questions of diversity: German contributions to benthos studies around South American and Antarctica, *Gayana*
   Gutt, J; Piepenburg, D (2003): Scale-dependent impacts of catastrophic disturbances by grounding icebergs on the diversity of Antarctic benthos, *Marine Ecology Progress Series*
   (and more)
   Size: unknown
   doi:10.1594/PANGAEA.198686 - Score: 80% - Similar datasets
2. **Gutt, J (2004):** Sea-bed photographs (benthos) from the Weddell Sea along ROV profile PS48/238 (©AWI, Gutt 1998)
   Reference: **Gutt, J; Arntz, WE; Balguerias, E et al. (2003):** Diverse approaches to questions of diversity: German contributions to benthos studies around South American and Antarctica, *Gayana*
   Gutt, J (2001): High latitude antarctic benthos: a coevolution of nature conservation and ecosystem research?, *Ocean and Polar Research*
   Gutt, J (2001): On the direct impact of ice on marine benthic communities, a review, *Polar Biology*
   (and more)
   Size: unknown
   doi:10.1594/PANGAEA.198685 - Score: 80% - Similar datasets
3. **Gutt, J (2004):** Sea-bed photographs (benthos) from the Weddell Sea along ROV profile PS48/219 (©AWI, Gutt 1998)
   Reference: **Gutt, J; Arntz, WE; Balguerias, E et al. (2003):** Diverse approaches to questions of diversity: German contributions to benthos studies around South American and Antarctica, *Gayana*
   Gutt, J; Piepenburg, 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*
   (and more)
   Size: unknown
   doi:10.1594/PANGAEA.198684 - Score: 80% - Similar datasets
4. **Gutt, J (2004):** Sea-bed photographs (benthos) from the Weddell Sea along ROV profile PS48/213 (©AWI, Gutt 1998)

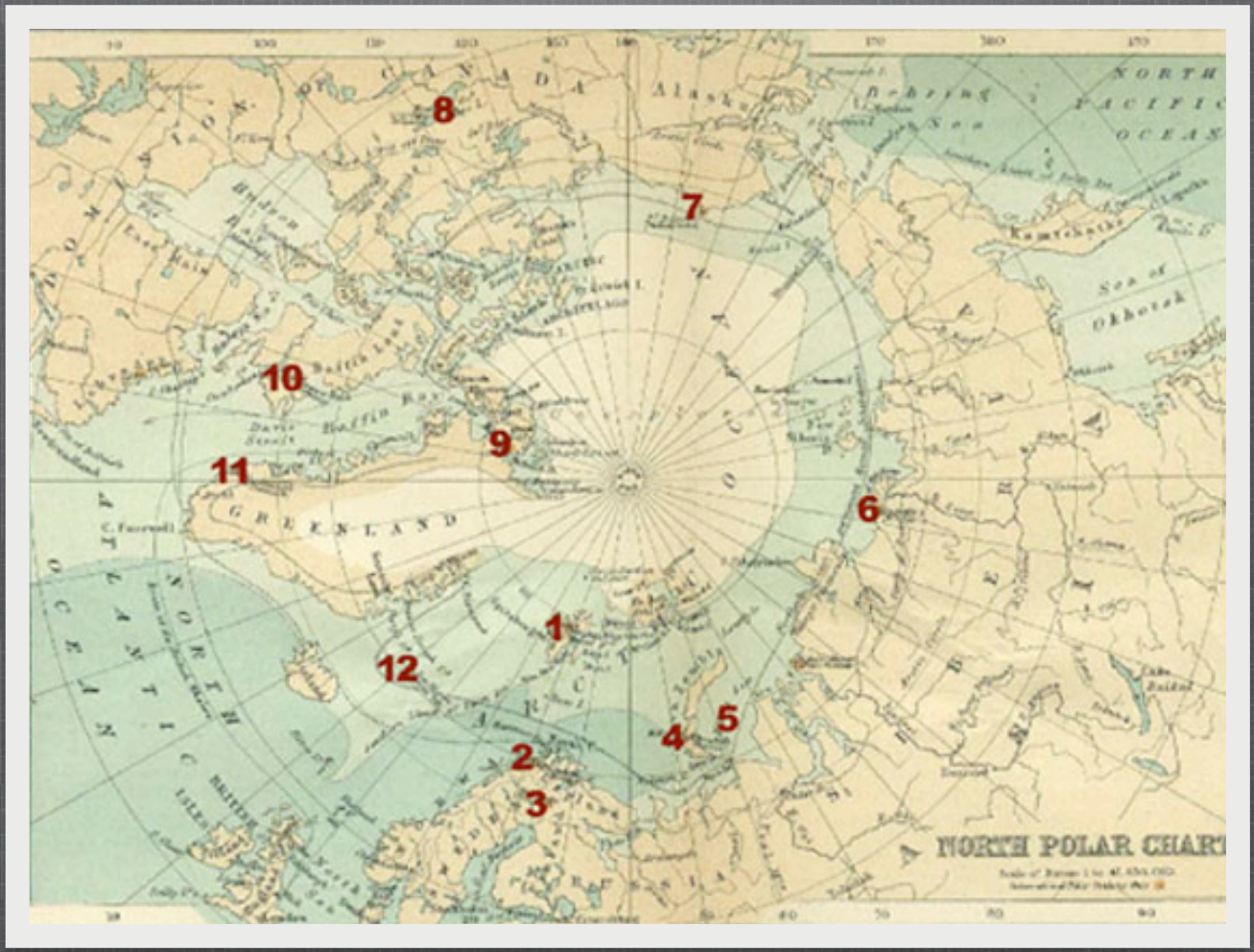
doi:10.1594/PANGAEA.319877

Sediment core documentation





doi:10.1594/PANGAEA.108079



International Polar Year (1882-1883)

Audio record
doi:10.1594/PANGAEA.339110

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