

https://dspace.org/

From Ions to Bits – Managing Data in a National Research Centre

Dr. Klump, Jens

2013-03-07
http://hdl.handle.net/10539/12502

Downloaded from a DSpace repository. https://dspace.org/

From Ions to Bits – Managing Data in a National Research Centre

Dr. Jens Klump
German Research Centre for Geosciences GFZ
jens.klump@gfz-potsdam.de

SADI Workshop on Digitisation and Digital Libraries: Standards, Best Practices, Policies and Technical Requirements

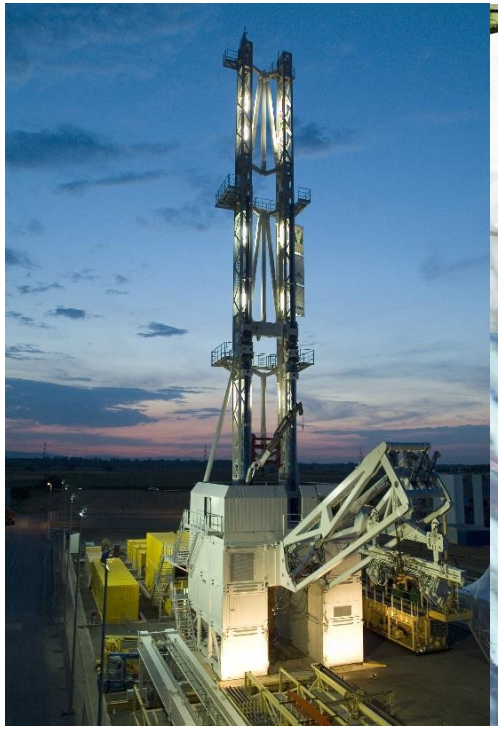
Johannesburg 2013-02-27

GFZ Helmholtz-Zentrum Potsdam

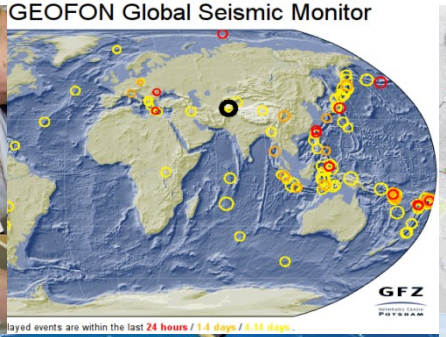
CC BY

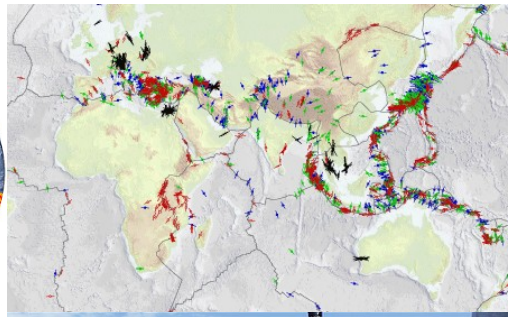
HELMHOLTZ GEMEINSCHAFT

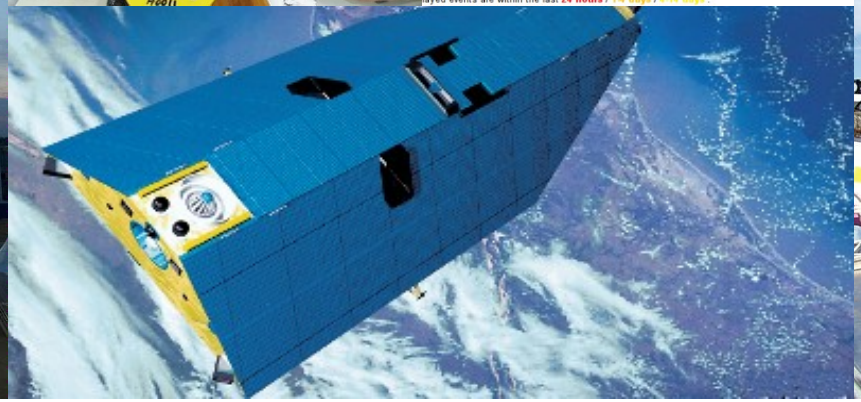
German Research Centre for Geosciences GFZ













Dealing with Research Data





... is like hearding cats.

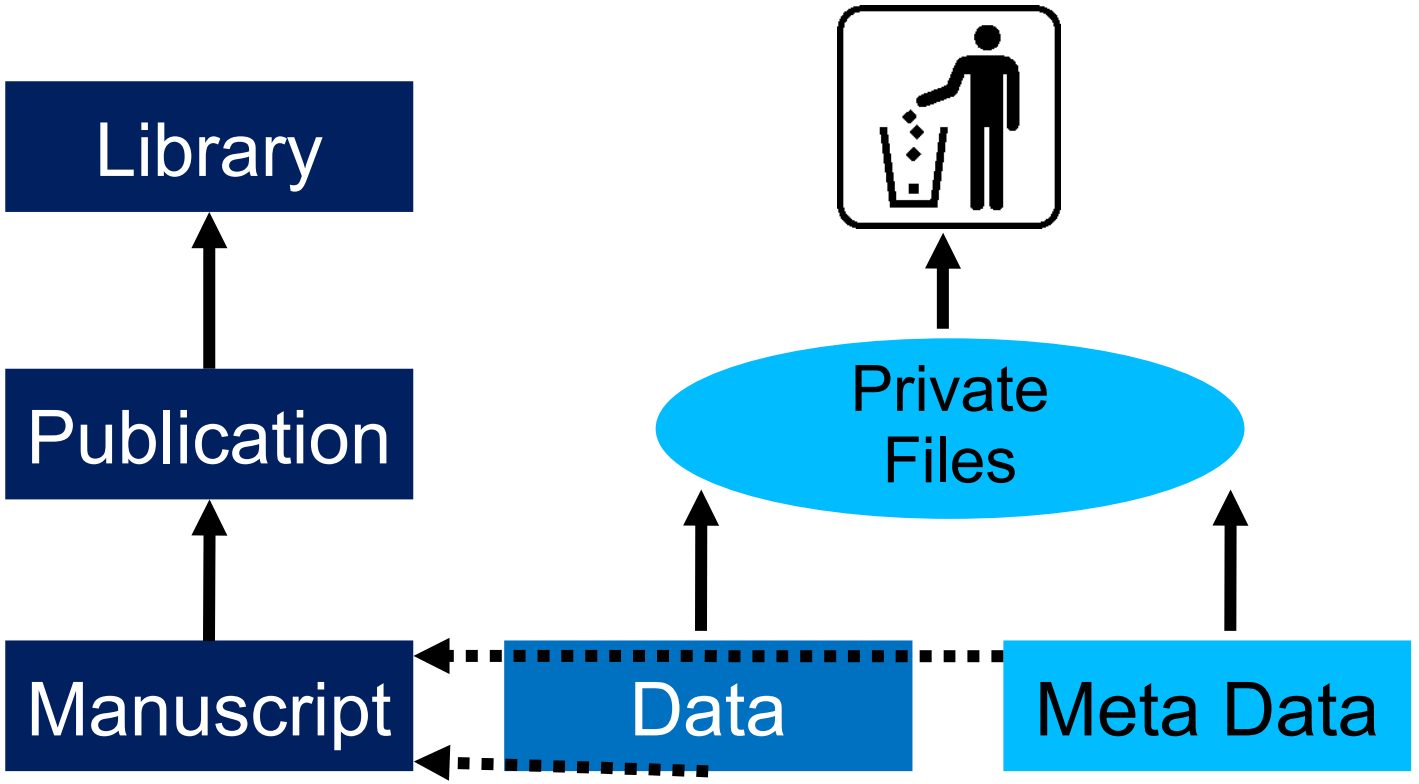
- Scope at GFZ:
  - Approx. 100 concurrent projects
  - Fluctuation of approx. 20 projects per year
  - Fluctuation of approx. 150 scientific staff per year
- We need generic tools to be able to cope!

Data Management Challenge

- The amount of data rises rapidly.
- Inaccessible data cannot be re-used.
- Consequences:
  - Duplication of efforts
  - Results cannot be verified
  - Data not available for reanalysis

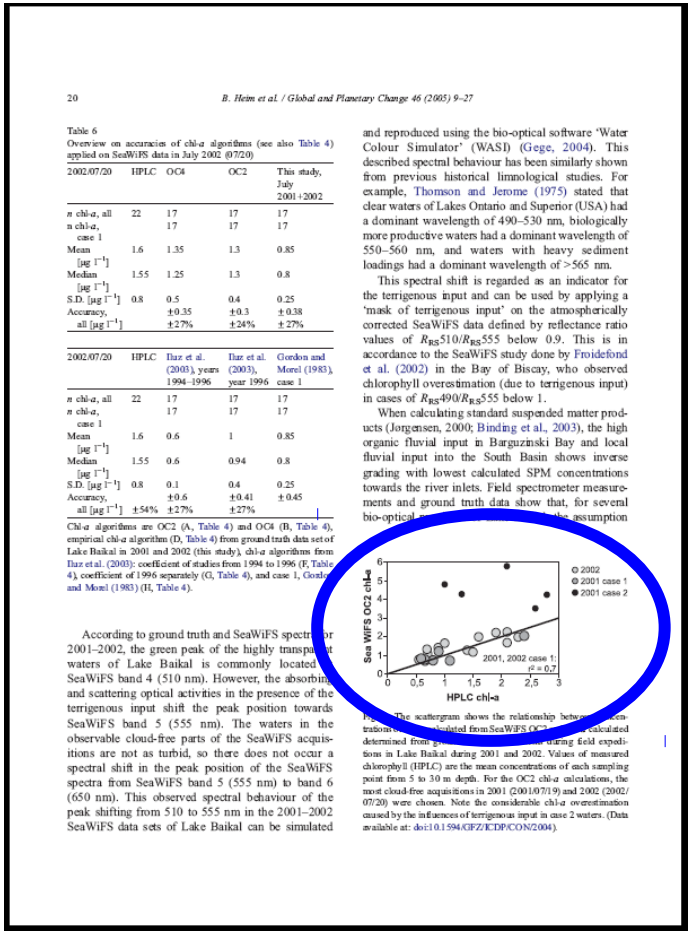


Dealing with Data: Status Quo



Helly *et al.* (2003)

Research Data Today



20 B. Heim et al. / Global and Planetary Change 46 (2005) 9–27

Table 6
Overview on accuracies of chl-*a* algorithms (see also Table 4) applied on SeaWiFS data in July 2002 (07/20)

2002/07/20	HPLC	OC4	OC2	This study, July 2001+2002
n chl-*a*, all	22	17	17	17
n chl-*a*, case 1		17	17	17
Mean [μg l^{-1}]	1.6	1.35	1.3	0.85
Median [μg l^{-1}]	1.55	1.25	1.3	0.8
S.D. [μg l^{-1}]	0.8	0.5	0.4	0.25
Accuracy, all [μg l^{-1}]		±0.35 ±27%	±0.3 ±24%	±0.38 ±27%

2002/07/20	HPLC	Ibaz et al. (2003), years 1994–1996	Ibaz et al. (2003), year 1996	Gordon and Morel (1983), case 1
n chl-*a*, all	22	17	17	17
n chl-*a*, case 1		17	17	17
Mean [μg l^{-1}]	1.6	0.6	1	0.85
Median [μg l^{-1}]	1.55	0.6	0.94	0.8
S.D. [μg l^{-1}]	0.8	0.1	0.4	0.25
Accuracy, all [μg l^{-1}]	±54%	±0.6 ±27%	±0.41 ±27%	±0.45

Chl-*a* algorithms are OC2 (A, Table 4) and OC4 (B, Table 4), empirical chl-*a* algorithm (D, Table 4) from ground truth data set of Lake Baikal in 2001 and 2002 (this study), chl-*a* algorithms from Ibaz et al. (2003): coefficient of studies from 1994 to 1996 (F, Table 4), coefficient of 1996 separately (G, Table 4), and case 1, Gordon and Morel (1983) (H, Table 4).

According to ground truth and SeaWiFS spectra for 2001–2002, the green peak of the highly transparent waters of Lake Baikal is commonly located at SeaWiFS band 4 (510 nm). However, the absorbing and scattering optical activities in the presence of the terrigenous input shift the peak position towards SeaWiFS band 5 (555 nm). The waters in the observable cloud-free parts of the SeaWiFS acquisitions are not as turbid, so there does not occur a spectral shift in the peak position of the SeaWiFS spectra from SeaWiFS band 5 (555 nm) to band 6 (650 nm). This observed spectral behaviour of the peak shifting from 510 to 555 nm in the 2001–2002 SeaWiFS data sets of Lake Baikal can be simulated and reproduced using the bio-optical software 'Water Colour Simulator' (WASI) (Gege, 2004). This described spectral behaviour has been similarly shown from previous historical limnological studies. For example, Thomson and Jerome (1975) stated that clear waters of Lakes Ontario and Superior (USA) had a dominant wavelength of 490–530 nm, biologically more productive waters had a dominant wavelength of 550–560 nm, and waters with heavy sediment loadings had a dominant wavelength of >565 nm.

This spectral shift is regarded as an indicator for the terrigenous input and can be used by applying a 'mask of terrigenous input' on the atmospherically corrected SeaWiFS data defined by reflectance ratio values of $R_{RS}510/R_{RS}555$ below 0.9. This is in accordance to the SeaWiFS study done by Froidefond et al. (2002) in the Bay of Biscay, who observed chlorophyll overestimation (due to terrigenous input) in cases of $R_{RS}490/R_{RS}555$ below 1.

When calculating standard suspended matter products (Jørgensen, 2000; Binding et al., 2003), the high organic fluvial input in Barguzinski Bay and local fluvial input into the South Basin shows inverse grading with lowest calculated SPM concentrations towards the river inlets. Field spectrometer measurements and ground truth data show that, for several bio-optical [illegible] assumption

The scattergram shows the relationship between concentrations [illegible] calculated from SeaWiFS OC2 [illegible] calculated determined from [illegible] during field expeditions in Lake Baikal during 2001 and 2002. Values of measured chlorophyll (HPLC) are the mean concentrations of each sampling point from 5 to 30 m depth. For the OC2 chl-*a* calculations, the most cloud-free acquisitions in 2001 (2001/07/19) and 2002 (2002/07/20) were chosen. Note the considerable chl-*a* overestimation caused by the influences of terrigenous input in case 2 waters. (Data available at: doi:10.1594/GFZ/ICDP/CON/2004).

Fig. 2. The scattergram shows the relationship between concentrations of chl-*a* calculated from SeaWiFS OC2 and chl-*a* calculated determined from ground truth measurements during field expeditions in Lake Baikal during 2001 and 2002. Values of measured chlorophyll (HPLC) are the mean concentrations of each sampling point from 5 to 30 m depth. For the OC2 chl-*a* calculations, the most cloud-free acquisitions in 2001 (2001/07/19) and 2002 (2002/07/20) were chosen. Note the considerable chl-*a* overestimation caused by the influences of terrigenous input in case 2 waters.

Data as Supplement to Literature

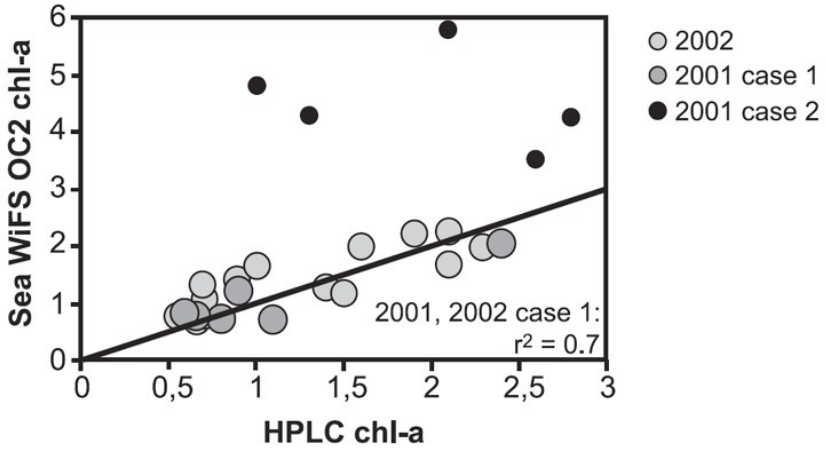


Fig. 2. The scattergram shows the relationship between concentrations of chl-*a* calculated from SeaWiFS OC2 and chl-*a* calculated determined from ground truth measurements during field expeditions in Lake Baikal during 2001 and 2002. Values of measured chlorophyll (HPLC) are the mean concentrations of each sampling point from 5 to 30 m depth. For the OC2 chl-*a* calculations, the most cloud-free acquisitions in 2001 (2001/07/19) and 2002 (2002/07/20) were chosen. Note the considerable chl-*a* overestimation caused by the influences of terrigenous input in case 2 waters.

doi:10.1594/GFZ.SDDB.1043

Data as Supplement to Literature

GFZ Helmholtz Centre Potsdam

Helmholtz Centre Potsdam
GFZ German Research Centre for Geosciences



HELMHOLTZ ASSOCIATION

Dataset Description

Search Datasets

Cite as Fietz, Susanne; Heim, Birgit; Oberhänsli, Hedi; Kaufmann, Hermann (2006): The relationship between concentrations of chl-a calculated from SeaWiFS OC2 and chl-a calculated determined from ground truth measurements during field expeditions in Lake Baikal during 2001 and 2002. Deutsches GeoForschungsZentrum GFZ. http://dx.doi.org/10.1594/GFZ.SDDB.1043

Abstract Values of measured chlorophyll (HPLC=High Pressure Liquid Chromatography) are the mean concentrations of each sampling point from 5 to 30 m depth. For the OC2 chl-a calculations, the least clouded acquisitions in 2001 (2001/07/19) and 2002 (2002/07/20) were chosen. Note the considerable chl-a overestimation caused by the influences of terrigenous input in case 2 waters.

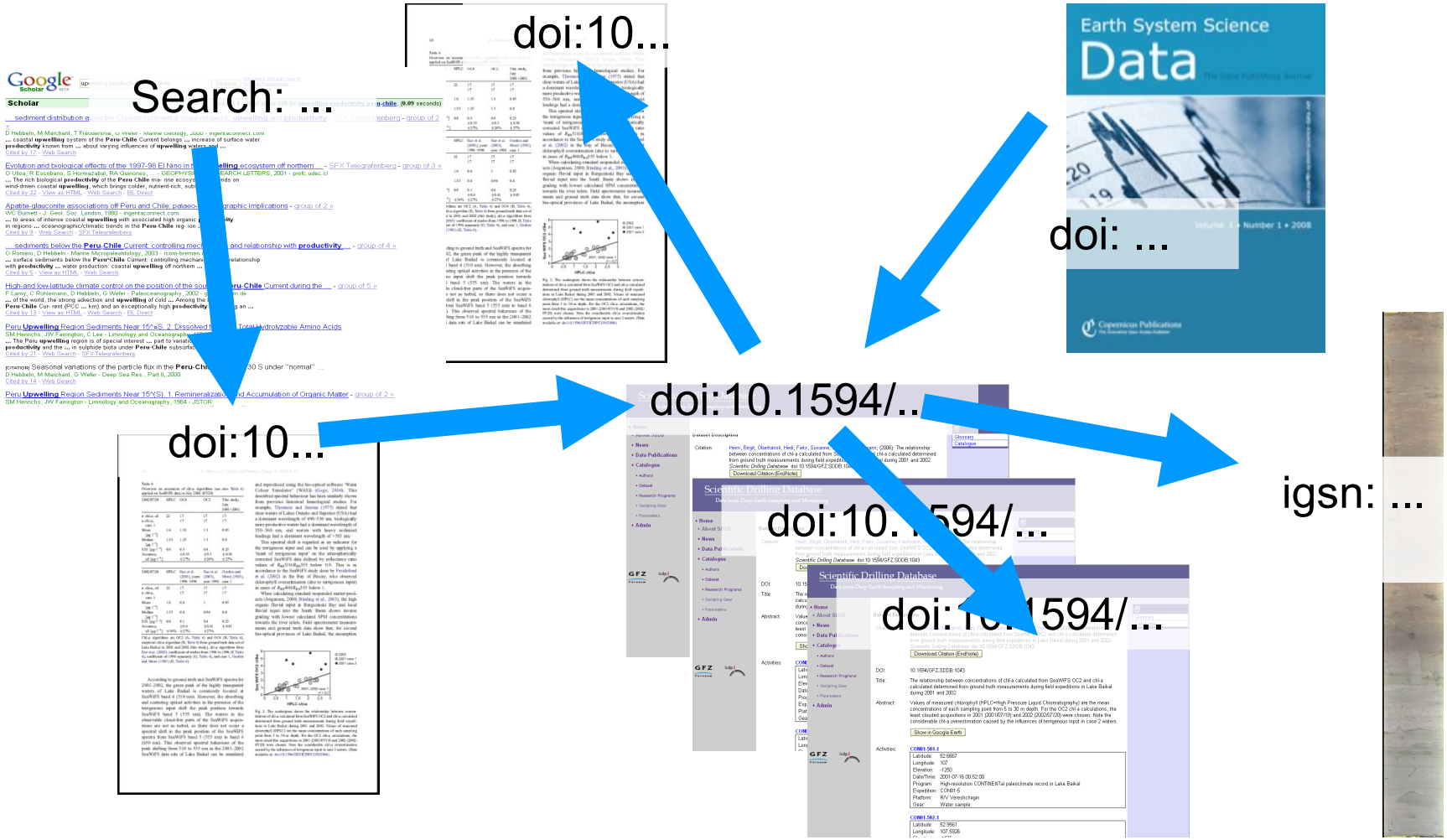
Supplement to Birgit Heim, Hedi Oberhaensli, Susanne Fietz and Hermann Kaufmann, Variation in Lake Baikal's phytoplankton distribution and fluvial input assessed by SeaWiFS satellite data, Global and Planetary Change, Volume 46, Issues 1-4, Progress towards reconstruct (http://dx.doi.org/10.1016/j.gloplacha.2004.11.011)

Location Latitude: 52.6667 Longitude: 107

Keywords Terrestrical Hydrosphere, Water Quality/Water Chemistry, Surface Water, HPLC chl-a concentration, OC2 chlorophyll-a concentration



Literature, Data, Samples



Linking Literature and Data

ScienceDirect - Marine Micropal...

Home | Browse | Search | My settings | My alerts | Help

Articles / Images — All fields — Journal/Book title — Author — Volume — Issue — Page — Search ScienceDirect — Search tip — Advanced search

Export citation | E-mail article

Abstract | Figures/Tables (13)

Marine Micropaleontology
Volume 66, Issues 3-4, 20 February 2008, Pages 208-221

doi:10.1016/j.marmicro.2007.10.002 | How to Cite or Link Using DOI

Permissions & Reprints

Centennial-scale climate variability in the Timor Sea during Marine Isotope Stage 3

Anke Dürkop[a], **Ann Holbourn**[a], **Wolfgang Kuhnt**[a], **Rina Zuraida**[a, b], **Nils Andersen**[c] and **Pieter M. Grootes**[c]

[a]Institute of Geosciences, Christian-Albrechts-University, Ludewig-Meyn-Str. 10-14, D-24118 Kiel, Germany

[b]Leibniz-Institute of Marine Sciences, IFM-GEOMAR, Wischhofstr. 1-3, D-24148 Kiel, Germany

[c]Leibniz-Laboratory for Radiometric Dating and Stable Isotope Research, Christian-Albrechts-University, Max-Eyth-Str. 11 - 13, D-24118 Kiel, Germany

Received 4 June 2007; revised 1 October 2007; accepted 4 October 2007. Available online 18 October 2007.

Abstract

We present a high-resolution (~ 60–110 yr) multi-proxy record spanning Marine Isotope Stage 3 from IMAGES Core MD01-2378 (13°04.95′ S and 121°47.27′ E, 1783 m water depth), located in the Timor Sea, off NW Australia. Today, this area is influenced by the Intertropical Convergence Zone, which

PANGAEA® – Supplementary Data
Paleoclimate investiations on sediment core MD01-2378



Related Articles

- Climate variability and land-ocean interactions in the ... *Marine Micropaleontology*
- Tropical warming in the Timor Sea led deglacial Antarct... *Earth and Planetary Science Letters*
- Direct comparison of mitochondrial markers for the anal... *Fisheries Research*
- Sensitivity of the Australian summer monsoon to tilt an...

Empty Archives



“Why is science publishing highly profitable while data repositories are virtually empty when compared to the volume of published literature?” (Ron Dekker, NWO)

Roles and Resposibilities

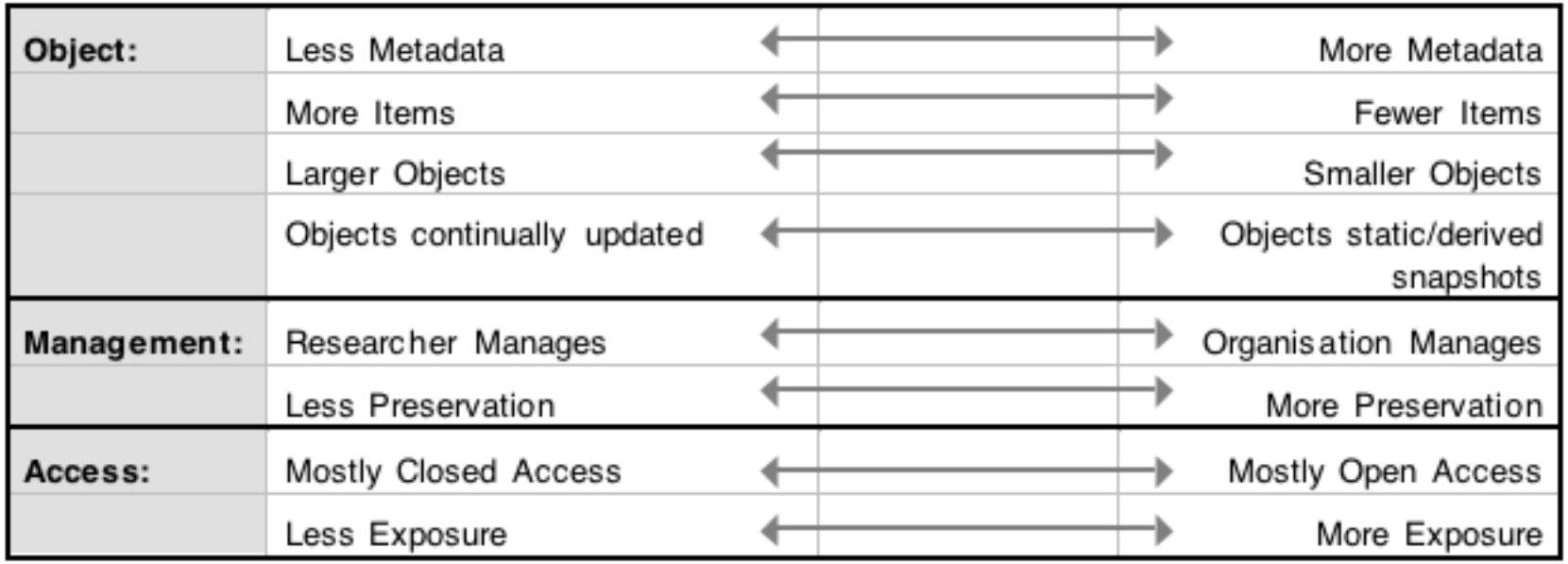
”I don't have time to spare for data management. For evaluation and tenure only my publications count.”

“People hate metadata.”

On the other hand: on eBay people describe physical objects with metadata by the thousands – and doing it right.

Have we assigned the wrong roles to our actors in the data life cycle?

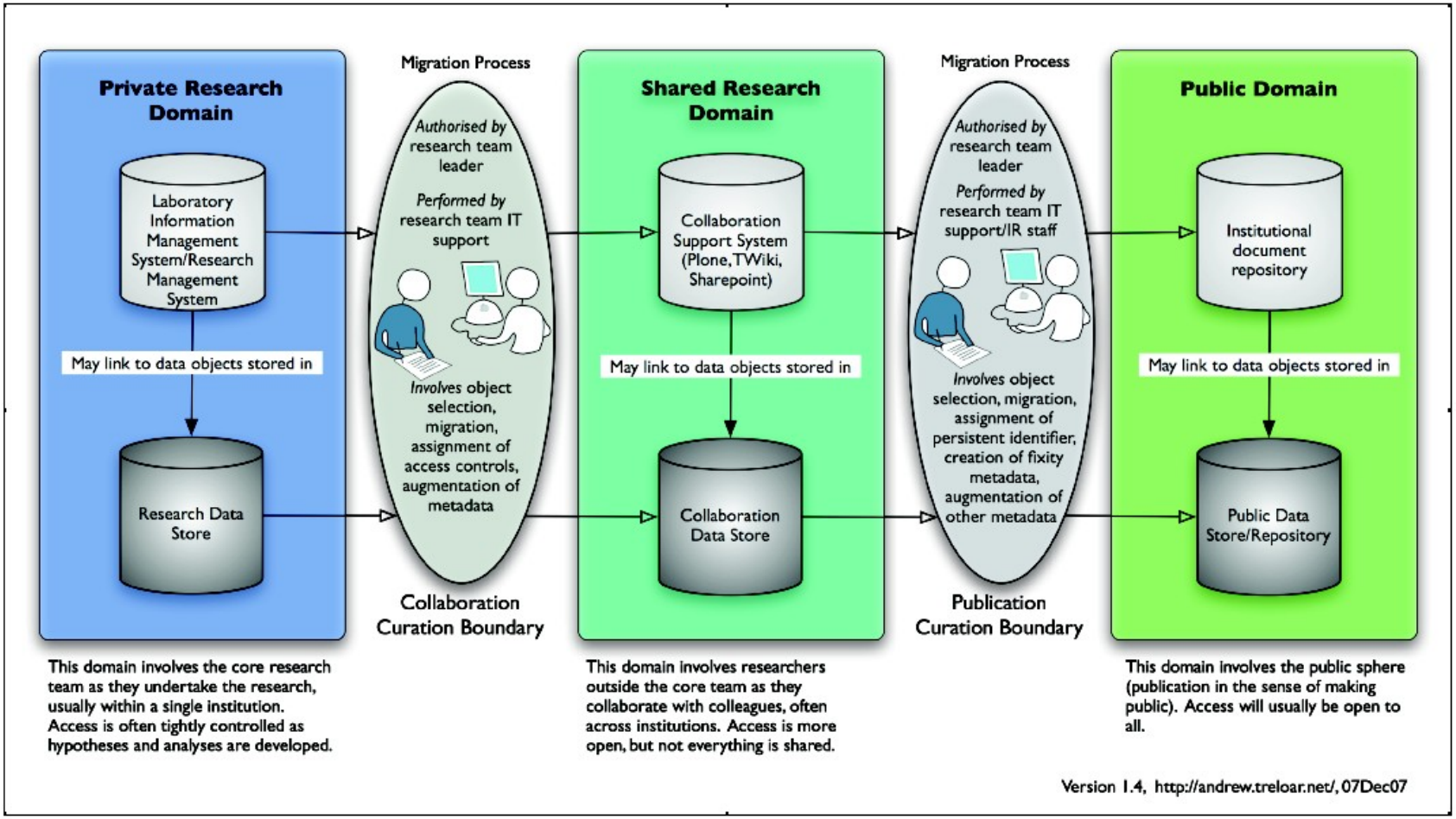
Data Curation Continuum



Object:	Less Metadata	⟷	More Metadata
	More Items	⟷	Fewer Items
	Larger Objects	⟷	Smaller Objects
	Objects continually updated	⟷	Objects static/derived snapshots
Management:	Researcher Manages	⟷	Organisation Manages
	Less Preservation	⟷	More Preservation
Access:	Mostly Closed Access	⟷	Mostly Open Access
	Less Exposure	⟷	More Exposure

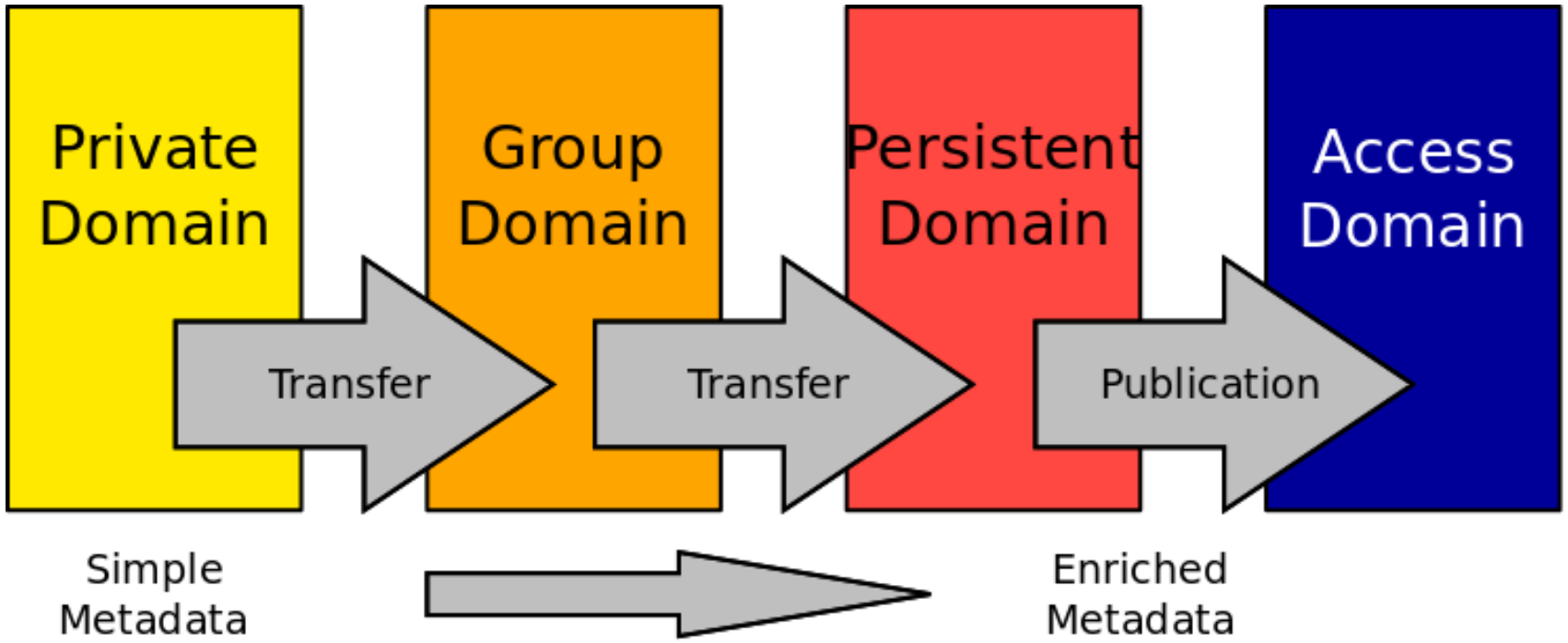
Treloar et al., 2007

Data Curation Continuum (2)



Treloar et al., 2008

Data Curation Continuum (3)



AG Forschungsdaten, 2010

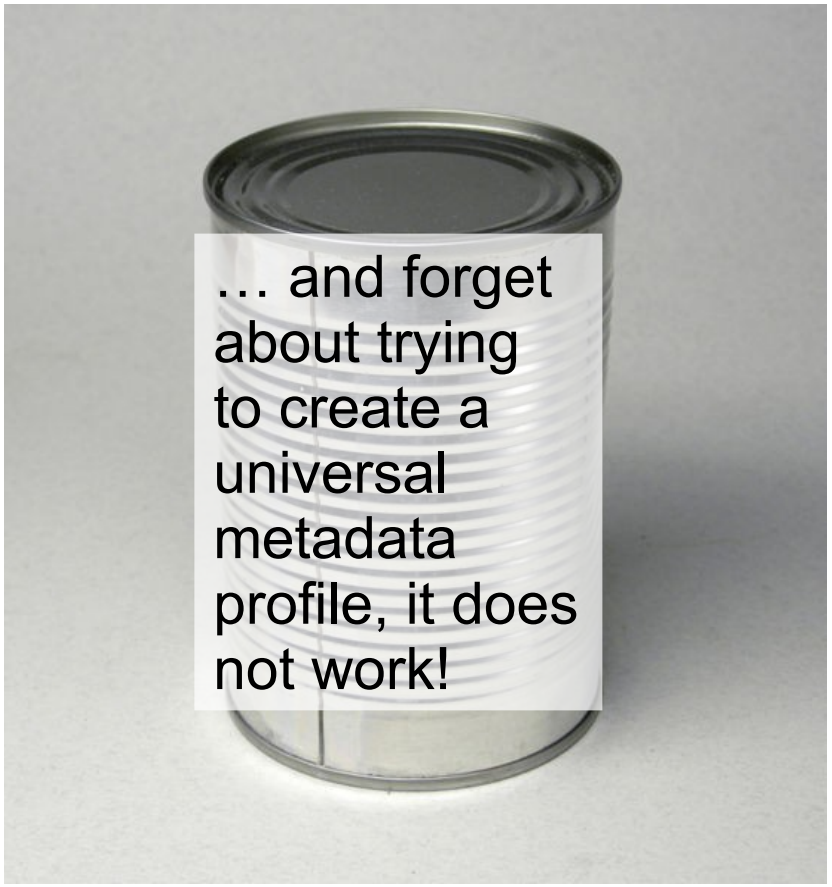
Describing Data

Metadata ... Oh – the pain!

Without description content cannot be discovered and re-used.

How do we do this without creating a "Data Bureaucracy"?

… by automation and by making it part of the workflow.

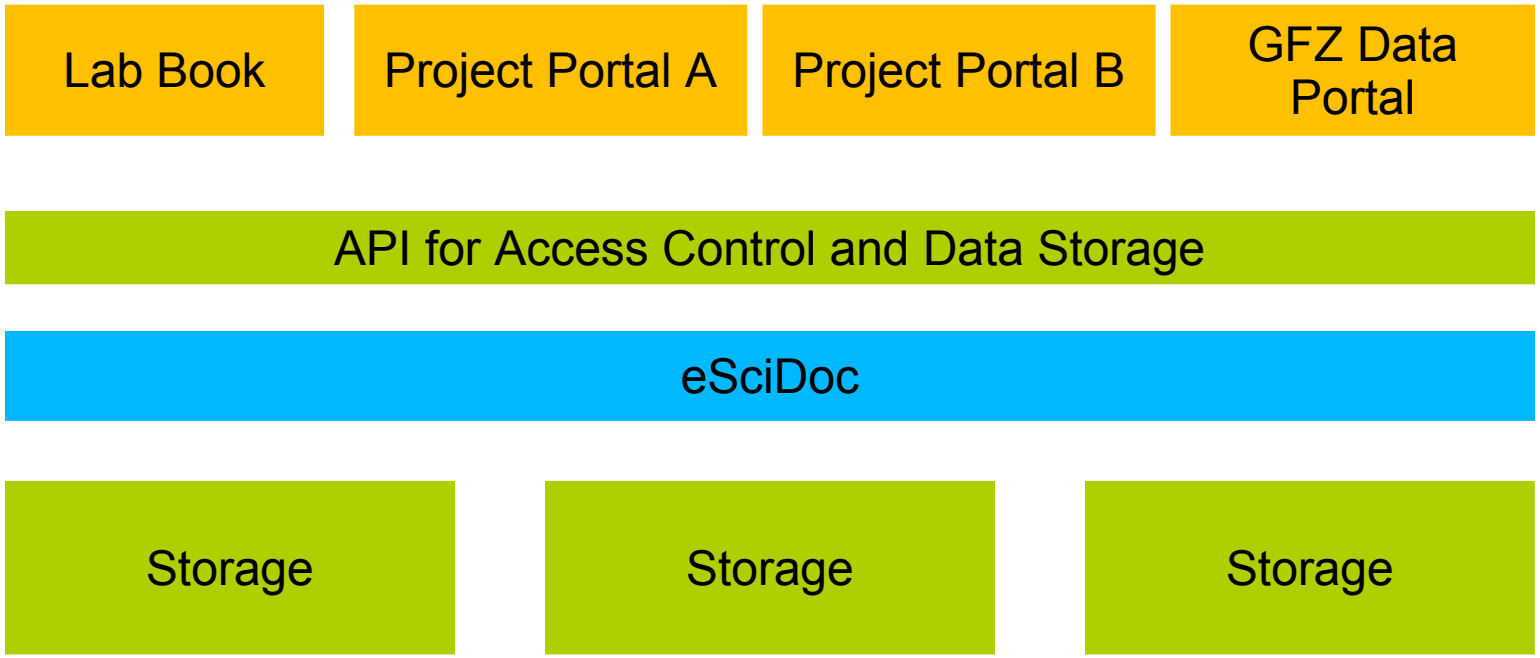


Institutional Workflows



- GFZ implemented a publications policy, a data policy is soon to follow.
- Publications DB is kept up to date by using only this DB for evaluation purposes.
- Workflow is supported by the library and the departmental offices.

Research Data Infrastructure



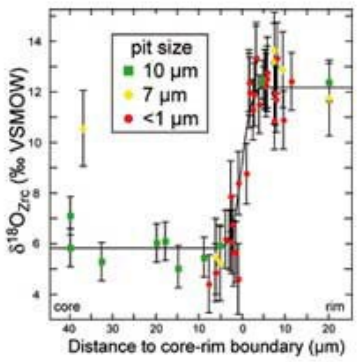
The research data infrastructure is available across all departments and projects at GFZ.

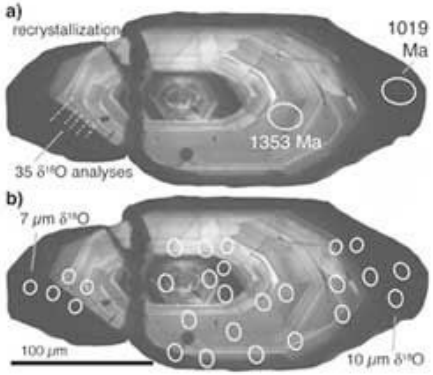
Shelf Storage for Data



From Ions to Bits







The GFZ Secondary Ion Mass Spectrometer (SIMS) is embedded in a virtual research environment:

- Lab Management
- Remote Control
- Data processing
- Data publication

Data are stored in and published through the GFZ data infrastructure.

Summary

Dealing with data has to be easy and attractive.

Current tools are not well integrated into the workflow of researchers.

Building “monolithic” data management tools is inflexible and does not scale.

In future we need more *prét à porter*, customizable tools, rather than *haute couture* boutique pieces.

Questions?



Thank you for your attention!