

Forschungsdatenmanagement in den Geowissenschaften

Dr. Jens Klump
Helmholtz-Zentrum Potsdam
Deutsches GeoForschungsZentrum
jens.klump@gfz-potsdam.de

Forschungsdatenmanagement in den Geowissenschaften

Die Geowissenschaften betrachten die feste Erde und die Prozesse an ihrer Oberfläche. Diese Vorgänge stehen gleichzeitig auch in Wechselwirkung mit Vorgängen in der Biosphäre, Atmosphäre, Hydrosphäre und Kryosphäre. Die räumlichen Dimensionen reichen über mehrere Größenordnungen von 10^{-10} m (Atome) bis 10^{12} m (Sonnensystem), die zeitlichen Dimensionen reichen von 10^{-15} s (molekulare Wechselwirkungen) bis 10^{17} s (Alter des Sonnensystems).



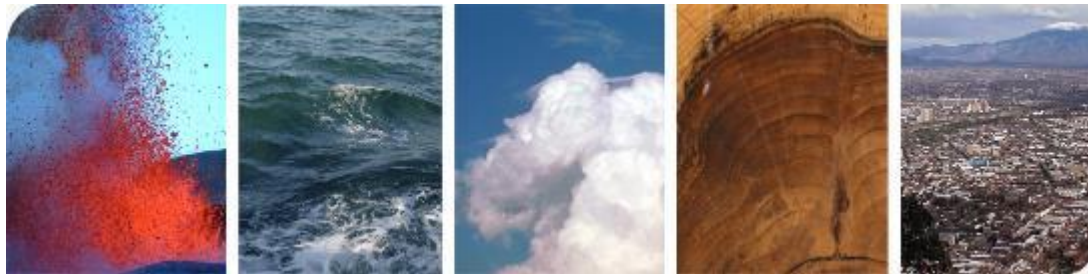
Herausforderung Datenmanagement

- Die Mengen der Forschungsdaten steigt rasant.
- Die unzugänglichen Daten werden unzureichend genutzt.
- Die Folgen:
 - Unnötige Doppelarbeiten
 - Forschungsergebnisse sind kaum verifizierbar
 - Widerspruch zur Guten wissenschaftlichen Praxis



Forschungsbereiche im Deutschen GeoForschungsZentrum

- Das Deutsche GeoForschungsZentrum ist in fünf Bereiche (Departments) gegliedert:
 - Geodäsie und Fernerkundung
 - Physik der Erde
 - Geodynamik und Geomaterialien
 - Chemie und Stoffkreisläufe der Erde
 - Prozesse der Erdoberfläche
- Jeder Bereich ist wiederum in Sektionen (insgesamt 25) gegliedert, jede mit einer Reihe von Projekten.
- Geschätzte Datenquellen > 150 verschiedene.



Beispiel GFZ Potsdam

Aus der Forschung am Helmholtz-Zentrum Potsdam Deutsches GeoForschungsZentrum entstehen eine Vielzahl von Datensätzen:

- **Erdbeobachtungssysteme** erzeugen meist große Datenmengen in teilweise automatisierten Workflows und mit weitgehend homogenen Strukturen.
- **Erdsystemmodellierung** erzeugt mäßig große Datenmengen mit eher homogenen Strukturen, aber meist ohne unterstützende Workflows.
- **Labordaten und Feldbeobachtungen** sind meist nur kleine Datenmengen, aber mit heterogenen Strukturen. Die daraus resultierenden Datensätze entstehen in ad-hoc formulierten Strukturen.

Policy am GFZ

- DFG Regeln für eine gute wiss. Praxis (GwP) wurden in eine Dienstanweisung überführt.
- Helmholtz-Gemeinschaft unterstützt den Offenen Zugang (Erstunterzeichner der Berliner Erklärung).
- GFZ Publikationordnung schließt die Regeln GwP und die Berliner Erklärung mit ein und ist Bestandteil des Arbeitsvertrags.

- Fazit: Daten müssten verfügbar sein, denn GwP und OA sind Bestandteile des Arbeitsvertrags, aber ...

Der Geist ist willig, ...

Wie lässt sich ein besserer Umgang mit Daten etablieren?

- Sanktionen greifen in diesem Umfeld nicht. Gibt es wirksame Anreize?
- Für die Umsetzung der Policies fehlen wichtige Bereiche:
 - Organisation und Integration in die wissenschaftlichen Arbeitsabläufe
 - Werkzeuge mit Integration in die wissenschaftlichen Arbeitsabläufe
- Heute ist der Umgang mit Daten oft nur Zweit- oder Drittaufgabe. Für den Einzelnen lohnt sich die investierte Zeit nicht.

Von Böcken und Gärtnern

„Für Datenmanagement habe ich keine Zeit, denn letztlich zählt bei der Begutachtung nur die Anzahl der Veröffentlichungen.“

„Forscher hassen Metadaten.“

Andererseits: in eBay beschreiben jeden Tag tausende von Menschen Objekte mit Metadaten – und machen es (weitgehend) richtig.

Sind möglicherweise die Rollen unserer Akteure im Lebenszyklus von Forschungsdaten falsch verteilt?

Umgang mit Forschungsdaten



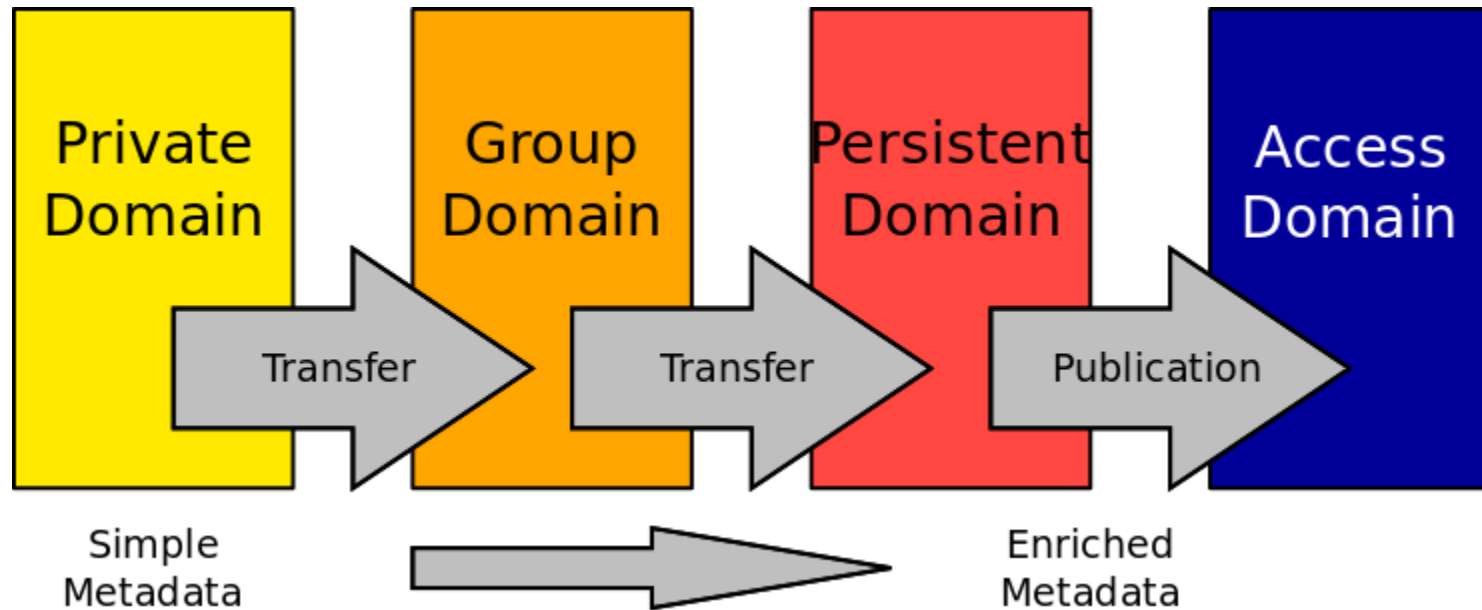
- ... ist wie eine Herde Katzen über die Prärie zu treiben.
- Umfang am GFZ:
 - Ca. 150 Projekte gleichzeitig
 - Fluktuation von ca. 30 Projekten pro Jahr
 - Fluktuation von ca. 200 Mitarbeitern pro Jahr
- Wir brauchen generische Werkzeuge!

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



Maßnahmen am GFZ

Infrastruktur für die systematische Veröffentlichung von Forschungsdaten mittels Digital Object Identifier (DOI).

Aufbau Virtueller Forschungsumgebungen für die Integration von Datenmanagement in die wissenschaftlichen Arbeitsabläufe.

Aufbau eines Repositoriums für Forschungsdaten außerhalb der bereits am GFZ bestehenden Systeme.

Arbeitsteilung zwischen CeGIT (Beratung), Rechenzentrum (IT-Infrastruktur) und Bibliothek (Veröffentlichung).

Institutionelle Workflows

Meldung

Publikations-DB

IR

Daten-Rep.

- GFZ hat die notwendigen Policy-Bausteine.
- Heute zählt bei der Evaluierung nur, was in der Literatur-DB steht.
- Workflow wird durch Bibliothek und Sekretariate unterstützt.
- Bibliothek bietet Plattform für die Veröffentlichung von Daten.

Publikation und Zitierbarkeit von Daten

DFG-Projekte STD-DOI und KOMFOR

Forschungsdaten in der Praxis

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
<i>n</i> chl- <i>a</i> , all	22	17	17	17
<i>n</i> chl- <i>a</i> , case 1	17	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	±0.3	±0.38	±0.38
Accuracy, all [µg l ⁻¹]	±27%	±27%	±24%	±27%

2002/07/20	HPLC	Itaz et al. (2003), years 1994–1996	Itaz et al. (2003), year 1996 case 1	Gordon and Morel (1983), case 1
<i>n</i> chl- <i>a</i> , all	22	17	17	17
<i>n</i> chl- <i>a</i> , case 1	17	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]	±0.6	±0.41	±0.45	±0.45
Accuracy, all [µg l ⁻¹]	±54%	±27%	±27%	±27%

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 Itaz et al. (2003); coefficient of studies from 1994 to 1996 (E, 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_{rs510}/R_{rs555} 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_{rs490}/R_{rs555} 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 parameters, the assumption

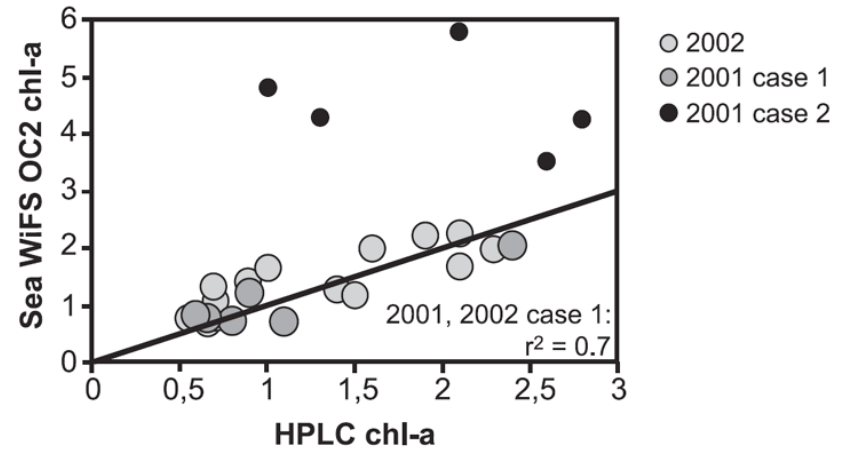


Fig. 2. The scattergram shows the relationship between concentrations of chl-*a* calculated from SeaWiFS OC2 and chl-*a* calculated determined from ground truth measurements during field expeditions in Lake Baikal during 2001 and 2002. Values of measured chlorophyll (HPLC) are the mean concentrations 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.

Scientific Drilling Database

Data from Deep Earth Sampling and Monitoring

- + Home
- + About SDDB
- + News
- + Data Publications

Dataset Description

Citation: [Heim, Birgit; Oberhänsli, Hedi; Fietz, Susanne; 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.](#)



Glossary
Catalogu

Citation: [Heim, Birgit; Oberhänsli, Hedi; Fietz, Susanne; 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. *Scientific Drilling Database*. doi:10.1594/GFZ.SDDB.1043](#)

[Download Citation \(EndNote\)](#)

Related Publications:

- Abstract: Values of measured chlorophyll (chl-a) by High-Pressure Liquid Chromatography on the moss
- [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 doi:10.1016/j.gloplacha.2004.11.011](#)



Related Publications:

- [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 doi:10.1016/j.gloplacha.2004.11.011](#)

Activities: [CON01-501-1](#)

Latitude: 52.6667 °N

Vernetzung von Literatur und Daten

ScienceDirect - Marine Micropal... +

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

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

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

^aInstitute of Geosciences, Christian-Albrechts-University, Ludewig-Meyn-Str. 10-14, D-24118 Kiel, Germany
^bLeibniz-Institute of Marine Sciences, IFM-GEOMAR, Wischhofstr. 1-3, D-24148 Kiel, Germany
^cLeibniz-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 Inter-tropical Convergence Zone, which

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



POWERED BY Google
Imagery ©2011, Map data ©2011 - Terms of Use

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...

Der Umgang mit Forschungsdaten soll einfacher werden

PanMetaDocs

Beschreibung der Daten

- Metadaten ... Oh – the pain!
- Ohne Beschreibung ist der Inhalt unbrauchbar.
- Wie kriege ich Metadaten ohne „Datenbürokratie“?
- Dokumentation muss einfacher werden!
- Beschreibungen der Daten sollten als Teil des Arbeitsablaufs (teil-) automatisch erfasst werden.



Werkzeuge

The screenshot displays the panMetaDocs web interface, divided into two main sections: 'Metadata' and 'File(s)'.
Metadata Section:
- **Preferences:** Entry form type: DataCite-2.0, Destination Folder: [dropdown], DOI: [dropdown], Do DOI operation: [button].
- **Identification:** Title: panMetaDocs - A tool for collecting and managing digital objects in a scientific research environment; Date: 2011-12-07; Author (1): Damian Ulbricht; Description: Data management in scientific projects is a challenging task. In many cases projects operate with a limited budget for data management that does not allow the development of customized software for data curation. On an institutional scale research data in the earth sciences are described by a number of different metadata schemas: panMetaWorks, which is the precursor to panMetaDocs, was developed to collect metadata and data in collaborative projects situated at more than one institution. Internet browsers allow easy interaction with panMetaWorks' PHP-based web user interface. Metadata are entered and data...; Project Name: [input].
- **NASA DIF Data:** Southernmost Latitude: 52.3795; Northernmost Latitude: 52.3795; Westernmost Longitude: 13.0648; Easternmost Longitude: 13.0648; Category: EARTH SCIENCE; Topic: Oceans; Term/Keyword (1): panMetaDocs; ISO Topic: Geoscientific Information; Data Center Contact: ulbricht@gfz-potsdam.de.
- **Publication:** DOI: 10.5072/GFZ.PMD.panMetaDocs; Resource Type: Dataset; Publisher (1): Deutsches GeoForschungszentrum; Publisher Short: GFZ Potsdam; Publication Year: 2011; language: en; copyright: [text area].
File(s) Section:
- **File upload (max. 100 Mb per file):** Please choose the file you wish to upload (txt, html usw.): [input]; Upload: [button]; WARNING: Do not distribute copyrighted material for which you don't own the rights or have permission from the owner.
- **Dataset versions:** 2012-01-02 12:25:30 (current); Change: [button].
- **Dataset Files:** PanMetaDocs-EGU.pdf (2 Mb); [button].

- Daten in Repositorien zu überführen ist immer noch viel zu kompliziert.
- Automatische Gewinnung von Metadaten:
 - aus dem Objekt
 - aus dem Kontext
- Werkzeuge müssen in neuen Projekten wiederverwendbar sein.

Forschungsdaten Infrastruktur

Laborbuch

Projekt-Portal

Daten-Portal

Einheitliche API für Zugriffsrechte und Datenverwaltung

eSciDoc

Speicher

Speicher

Speicher

eSciDoc als „Hochregal“ für Daten



Fernsteuerbares Sekundär-Ionen-Massenspektrometer

VirtualSIMS

VirtualSIMS

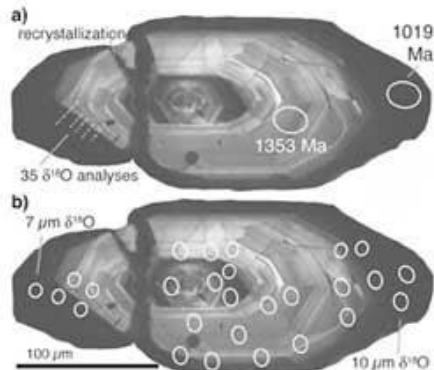
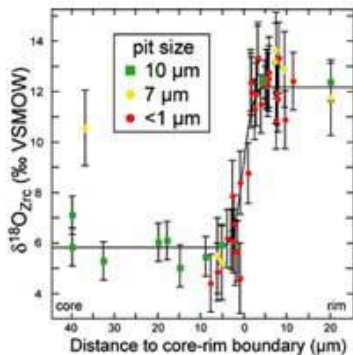


SIMS ist ein extrem leistungsfähiges Verfahren zur Bestimmung von Isotopenverhältnissen.

Weltweit gibt es derzeit nur 50 Geräte diesen Typs, für die meisten Wissenschaftler sind sie nicht zugänglich.

Über eine virtuelle Arbeitsumgebung soll das Gerät einem größeren Kreis von Nutzern weltweit zugänglich gemacht werden.

Fernsteuerung erlaubt 24-Stunden Betrieb.



Zusammenfassung und Ausblick

Das GFZ misst Forschungsdaten eine große Bedeutung bei. Der Policy-Rahmen für den Umgang mit Daten am GFZ lässt kaum zu wünschen übrig:

Lücken in der Umsetzung:

- Werkzeuge
- Integration in die Arbeitsabläufe der Forschung
- Support-Strukturen für Small Science Data

Durch VRE soll Datenmanagement enger mit den Arbeitsabläufe in der Forschung verbunden werden.

Denkbar ist auch, Datenmanagement-Werkzeuge als Cloud-Dienste in Leistungszentren für Forschungsdaten anzubieten.

Fragen?



Danke für Ihre Aufmerksamkeit!