

Persistente Identifikatoren - Anwendungsfälle und Systeme für die Naturwissenschaften

Dr. Jens Klump

Helmholtz-Zentrum Potsdam
Deutsches GeoForschungsZentrum GFZ

Übersicht

- Was sind persistente Identifikatoren?
- Welche Systeme für persistente Identifikatoren gibt es?
- Wo werden persistente Identifikatoren eingesetzt?
- Was ist bei der Auswahl von Identifier-Systemen zu beachten?



Was war die Frage?

- Ressourcen im Internet (z.B. Dateien oder Dienste) werden allgemein über ihren Universal Resource Locator (URL) referenziert.
- Problem: Der URL bezeichnet nur den Ort eines Objekts.
- Die Halbwertszeit von URL beträgt ca. 3 Jahre.

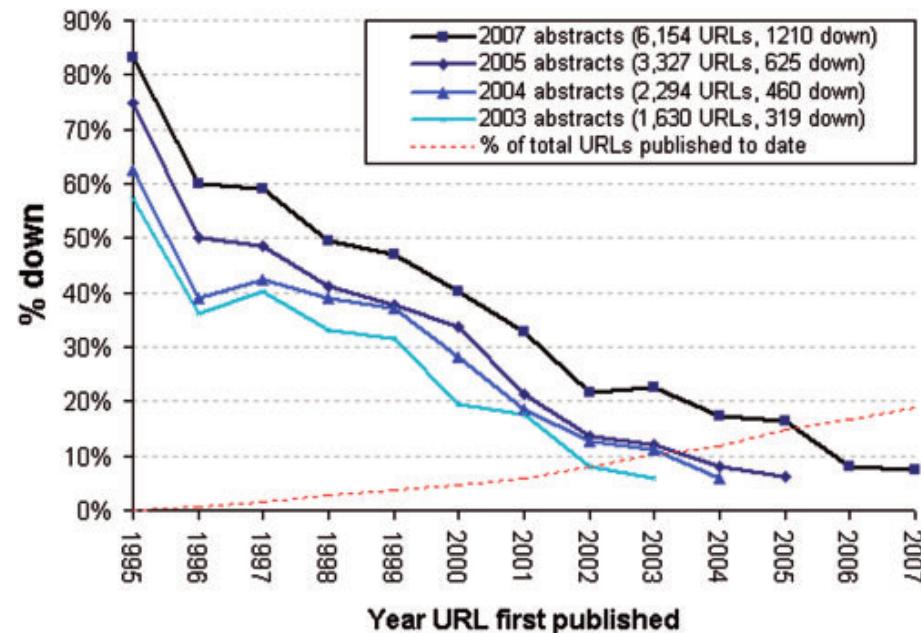


Abb.: Wren (2008)

Das Problem

- Wenn das Objekt an einen anderen Ort im Netz wechselt ist es nicht mehr ohne weiteres auffindbar.
- Forschung braucht eindeutige Verweise auf die Forschungsobjekte.

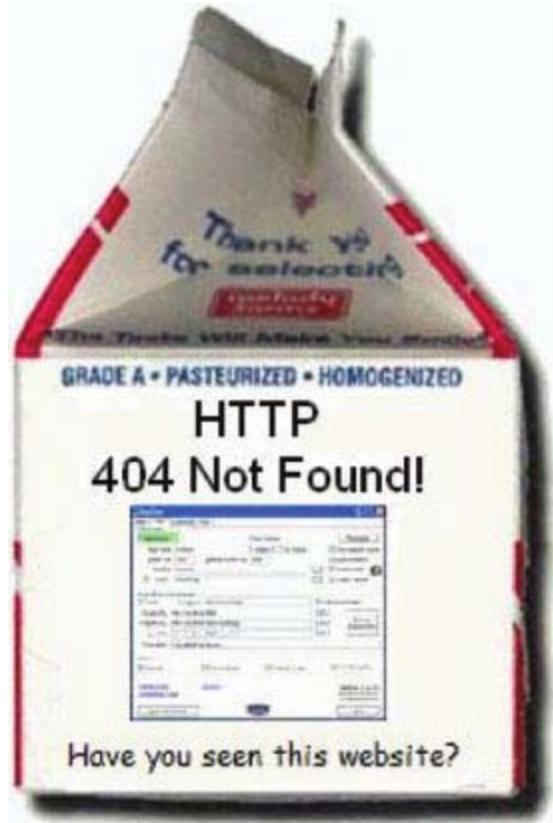


Abb.: Wren (2008)

Definition: Persistente Identifikatoren

- **Identifier:** Ein **Name** (name), der eindeutig mit einem **Objekt** (thing) verknüpft ist.
- **Name:** Ein Name wird durch eine Zeichenkette repräsentiert. Er gehört zu einem identifizierbaren Namenssystem (context).
- **Objekt** (thing): Ein Objekt kann alles sein, worüber gesprochen werden kann, insbesondere alles, was unterschieden und mit einem Identifier versehen werden kann. Es können also z.B. statische oder dynamische Objekte, Dokumente oder Prozeduren und auch aggregierte Objekte oder Teilobjekte identifiziert werden.
- **Persistenz:** Persistenz heißt, dass ein Identifier **auf Dauer eindeutig** mit einem Objekt verknüpft bleibt.



Von PID zu URL

- Um das benannte Objekt anhand seines Namens finden zu können braucht ein PID-System einen Auflösungsdienst, der den Namen des Objekts in einen URL übersetzt.
- Um einen PID nutzen zu können muss man den entsprechenden Auflösungsdienst kennen.
- Beispiel DOI:
 - doi:10.1000/182
 - <http://dx.doi.org/10.1000/182>
- Für die Wissenschaft von besonderem Interesse sind global auflösbare PID.



Beispiele

- Handle
 - DOI (Digital Object Identifier)
 - EPIC (European Persistent Identifier Consortium)
 - IGSN (International Geo Sample Number)
 - Andere Handle-Dienste
- URN
 - nbn-urn (Nationalbibliothek)
 - LSID (Life Science Identifiers)
- Permalink und PURL
- ARK



Was wird referenziert?

- **Objekt** (thing): Ein Objekt kann alles sein, worüber gesprochen werden kann, insbesondere alles, was unterschieden und mit einem Identifier versehen werden kann. Es können also z.B. statische oder dynamische Objekte, Dokumente oder Prozeduren und auch aggregierte Objekte oder Teilobjekte identifiziert werden.
- Im Internet referenziert werden können nur digitale Objekte oder digitale Repräsentationen von Objekten.
 - Literatur
 - Autoren und Institutionen
 - Datensätze
 - Proben (Biologie, Geologie, Genetik, ...)



Benennung der Objekte

- Der Begriff Persistenz impliziert, dass die referenzierten Objekte für lange Zeit verfügbar sein sollen, ggf. über das Ende ihrer beheimatenden Organisation hinaus.
- Da der Rahmen um ein Objekt herum sich ändern kann, während das Objekt selbst unverändert bleibt, wird empfohlen, den Namen des Objekts möglichst nicht sprechend zu formulieren.
- Die Regeln für die Bildung eines Namens sollten klar und logisch sein und nach Möglichkeit aus bereits existierenden Identifikatoren abgeleitet werden.



Granularität

- Für die Granularität der Objekte gibt es keine allgemeine Regel.
- Ein Nutzer sollte über den PID eines Objekts zu einer nachnutzbaren Entität geleitet werden.
- Viele PID-Systeme können auch semantische Beziehungen zwischen Objekten abbilden.
- Beispiel: König-Langlo, G., und H. Gernandt (2009), 426 ozonesonde profiles from Georg-Forster-Station, 10.1594/PANGAEA.547983
 - DOI bezieht sich auf eine Sammlung von 426 Datensätzen
 - Jeder der 426 Datensätze hat einen eigenen DOI
- Auch nicht abgeschlossene Reihen sind möglich.



Organisation und Persistenz

- Die Persistenz und Vertrauenswürdigkeit eines PID-Systems kann anhand eines Kriterienkatalogs eingeschätzt werden:
 - Organisatorischer Rahmen (Trägerschaft, Verbindlichkeit, Geschäftsmodell, Rechte, Neutralität, Exit-Strategie, ...)
 - Umgang mit Objekten (Skalierbarkeit, Eindeutigkeit, Gültigkeit, Sicherheit, Transparenz, Einfachheit, ...)
 - Infrastruktur und Sicherheit (Standortunabhängigkeit, Sicherheit, Ausfalllösung, Technologieunabhängigkeit, Migrierbarkeit, Skalierbarkeit)
- Technisch ist die Einrichtung eines PID-Service keine große Herausforderung.



Anwendungen

20

B. Heim et al. / Global and Planetary Change 46 (2006) 9–27

Table 6
Overview on accuracy of chl-a algorithms (see also Table 4)
applied on SeaWiFS data in July 2002 (0720).

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 [$\mu\text{g l}^{-1}$] | 1.6 | 1.35 | 1.3 | 0.85 |
| Median [$\mu\text{g l}^{-1}$] | 1.55 | 1.25 | 1.3 | 0.8 |
| S.D. [$\mu\text{g l}^{-1}$] | 0.8 | 0.5 | 0.4 | 0.25 |
| Accuracy, all | ± 0.35 | ± 0.3 | ± 0.3 | ± 0.38 |
| | $\pm 2.7\%$ | $\pm 24\%$ | $\pm 27\%$ | |

2002/07/20 HPLC Diaz et al.
(2003), years
1994–1996 year 1996 case 1

| | | | | |
|------------------------------------|-------------|-------------|------------|------|
| n chl-a, all | 22 | 17 | 17 | 17 |
| n chl-a, case 1 | | 17 | 17 | 17 |
| Mean [$\mu\text{g l}^{-1}$] | 1.6 | 0.6 | 1 | 0.85 |
| Median [$\mu\text{g l}^{-1}$] | 1.55 | 0.6 | 0.94 | 0.8 |
| S.D. [$\mu\text{g l}^{-1}$] | 0.8 | 0.1 | 0.4 | 0.25 |
| Accuracy, all | ± 0.6 | ± 0.41 | ± 0.45 | |
| | $\pm 5.4\%$ | $\pm 2.7\%$ | $\pm 27\%$ | |

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 Diaz 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 field measurements. For example, Thompson and Jerome (1973) 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 R_{490}/R_{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_{490}/R_{555} below 1.

When calculating standard suspended matter products (Jørgensen, 2006; 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 provinces of Lake Baikal, the assumption

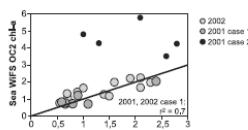


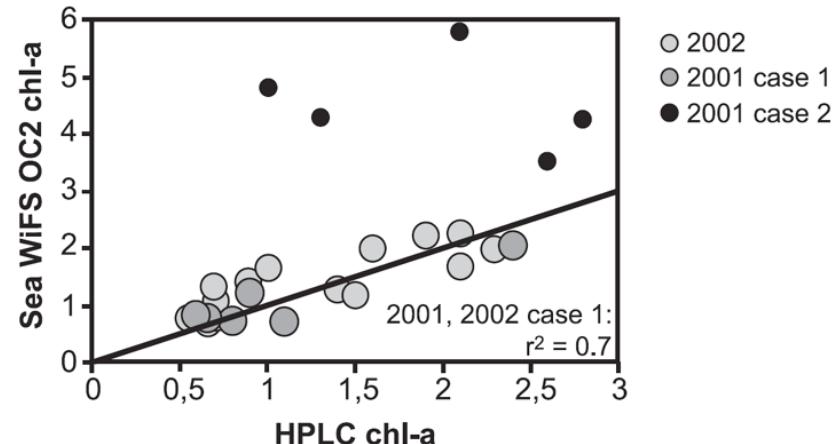
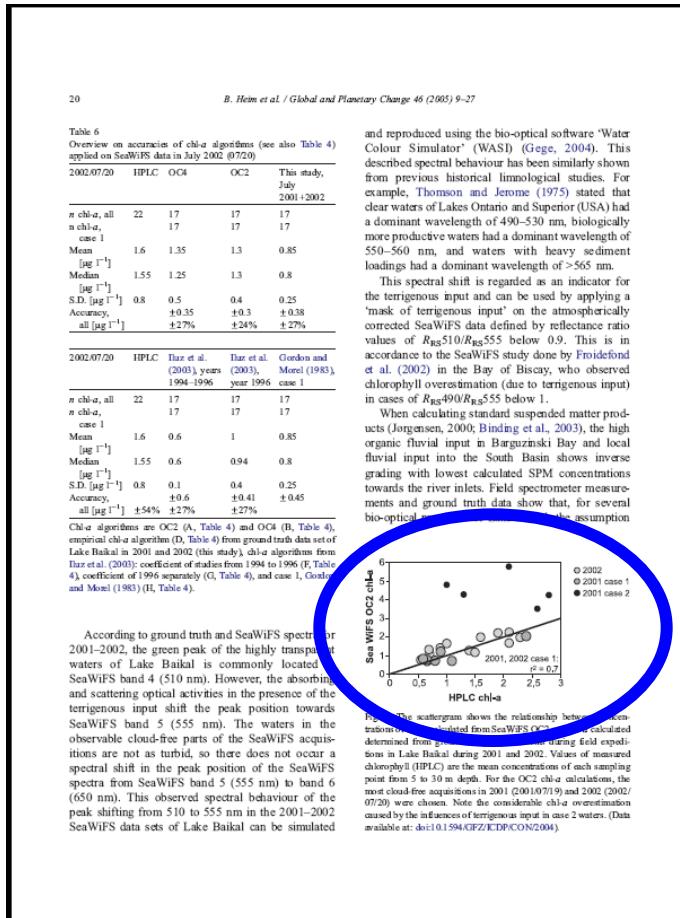
Fig. 2. The scattergram shows the relationship between concentrations of chl-a calculated from SeaWiFS OC2 and chl-a calculated determined from ground truth measurements during field expeditions in Lake Baikal during 2001 and 2002. Values of measured chlorophyll (HPLC) are the mean concentrations of each sampling point for a depth of 10 m depth. For the OC2 chl-a values, the most cloud-free acquisitions in 2001 (20010719) 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.)

- Am weitesten verbreitet finden sich PID heute für die Identifizierung von Fachartikeln, aber auch andere Anwendungen sind denkbar.
- Zum Einsatz kommen meistens DOI, URN und Permalinks.

DataCite

- DataCite ist eine internationale Organisation für die Vergabe von DOI für Datenveröffentlichungen.
- Mitglieder in DataCite sind überwiegend Nationalbibliotheken als Registrierungsagenten (RA) für Datenveröffentlichungen.
- RA in Deutschland sind TIB, ZB MED, GESIS und ZBW.
- Die RA delegieren Qualitätssicherung und Speicherung der Daten an Publikationsagenten.

Forschungsdaten in der Praxis

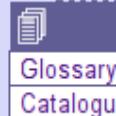


Scientific Drilling Database

Data from Deep Earth Sampling and Monitoring

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

Dataset Description



Glossary

Catalogue

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.](#)

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.SDDDB.1043

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

- + Admin

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](https://doi.org/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](https://doi.org/10.1016/j.gloplacha.2004.11.011)

Activities:

[CON01-501-1](#)

Latitude: 52.6667 °N

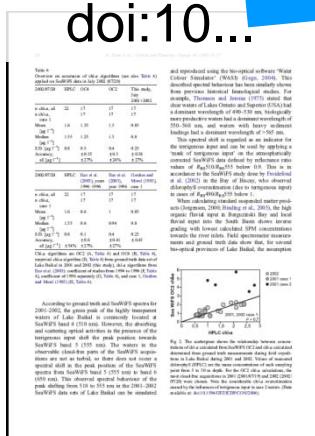
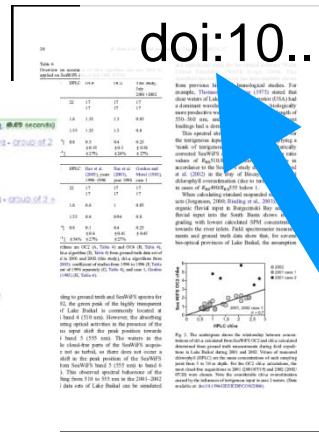
Literatur – Daten – Proben



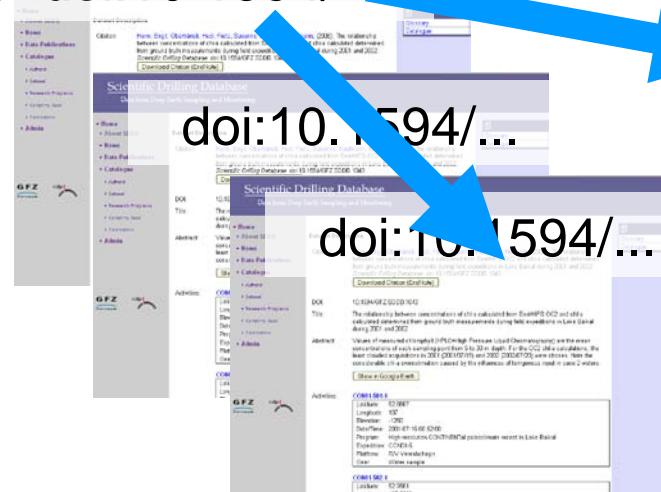
Scholar

...searched in

Suche: ...



doi:10.1594/...



igsn: ICDP...



Vernetzung von Literatur und Daten

ScienceDirect - Marine Micropal...

Home | Browse | Search | My settings | My alerts

Articles All fields Author
Images Journal/Book title Volume Issue Page

Advanced search | Help | Export citation | E-mail article | Search ScienceDirect | Search tip

Abstract | Figures/Tables (13)

Marine Micropaleontology
Volume 66, Issues 3-4, 20 February 2008, Pages 208-221
doi:10.1016/j.mamicro.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 (\sim 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 Intermonsoonal Convergence Zone, which

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

Hybrid | Export citation | E-mail article | Search PANGAEA | Help

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 Antarctic ... *Earth and Planetary Science Letters*
- Direct comparison of mitochondrial markers for the anal... *Fisheries Research*
- Sensitivity of the Australian summer monsoon to tilt an...



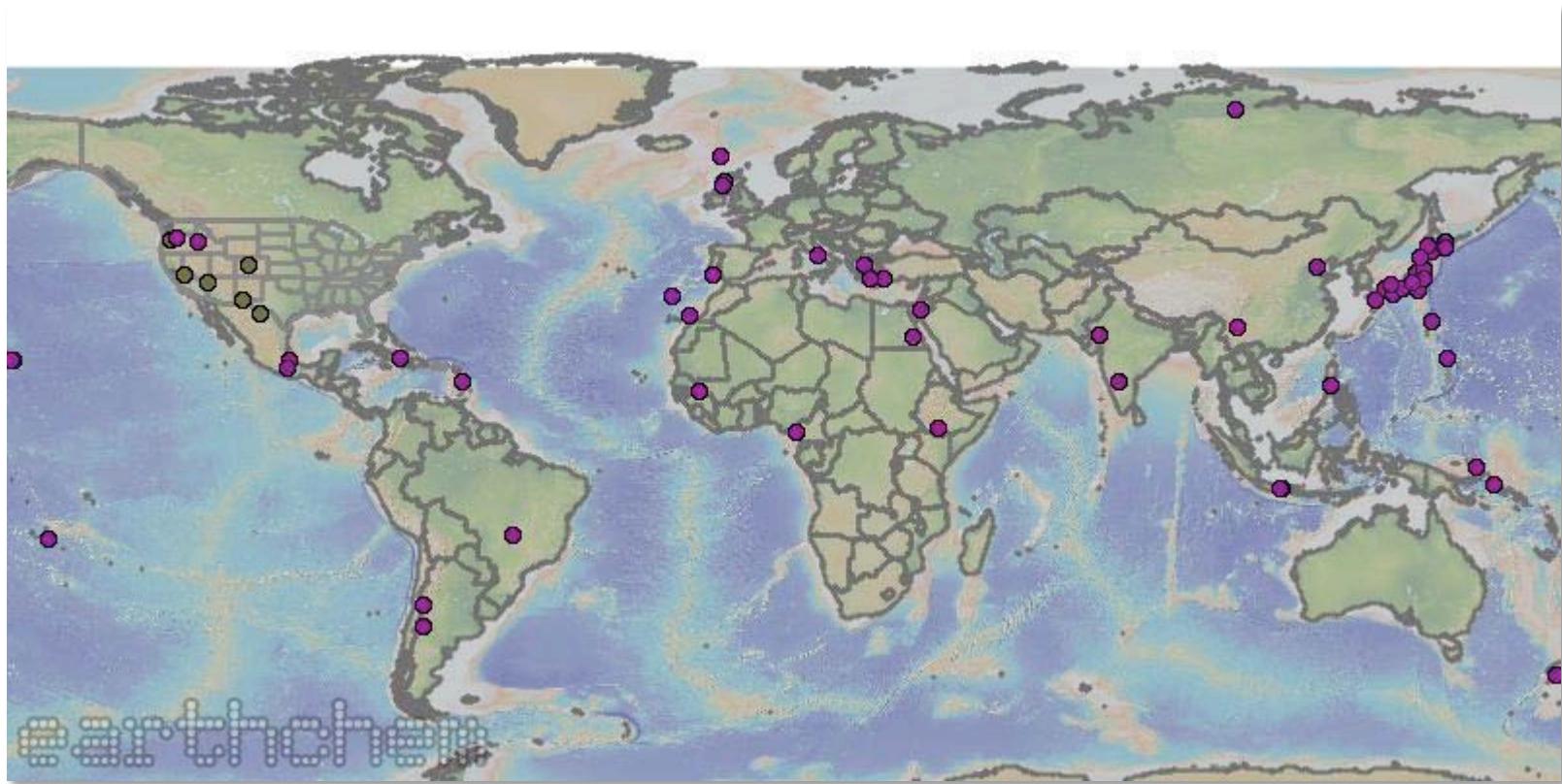
PID für physische Objekte

| | |
|--------------------------------|--|
| | IGSN: GMY00007W Sample Name: TN182_47_002 Other Name(s): Sample Type: Individual Sample Parent IGSN: GMY00001B |
| Description | |
| Material: | Rock |
| Classification: | Igneous>Plutonic>Mafic |
| Field Name: | gabbro, hornblende gabbro |
| Description: | mafic plutonic rock |
| Age (min): | 2 Ma |
| Age (max): | 50 Ma |
| Collection Method: | Dredging |
| Collection Method Description: | Not Provided |
| Size: | 0.5 kg |
| Geological Age: | Cenozoic |
| Geological Unit: | Not Provided |
| Comment: | Not Provided |
| Purpose: | Not Provided |
| Geolocation | |
| Latitude: | 51.6298 |
| Longitude: | 179.8611 |
| Elevation: | -1460 |
| Nav Type: | Not Provided |
| Physiographic Feature: | fault scarp |
| Name Of Physiographic Feature: | Sunday Basin |
| Location Description: | north wall of Sunday Basin depression |

- PID können auch für physische Objekte verwendet werden.
- Der PID verweist dann auf eine digitale Repräsentation des Objekts.



Anwendungen



Lokationen einer Probe „M1“ in der EarthChem Datenbank.



Beispiel: IGSN



Die International Geo Sample Number verweist auf eine Internetseite, auf der z.B. ein Bohrkern beschrieben wird.



Fragen?



Danke für Ihre Aufmerksamkeit!



Helmholtz Open Access Webinare zu Forschungsdaten
Webinar 1 - 22.06. und 04.07.2012

Literatur

- Bütikofer, N. (2009), Kriterienkatalog zur Prüfung der Vertrauenswürdigkeit von PI-Systemen, nestor-Materialien. <http://nbn-resolving.de/urn:nbn:de:0008-20080710140>
- Hakala, J. (2010), Persistent identifiers – an overview, KIM-TWR. <http://metadaten-twr.org/2010/10/13/persistent-identifiers-an-overview/>
- Simons, N. (2012), Implementing DOIs for Research Data, D-Lib, 18(5/6), doi:10.1045/may2012-simons
- Wren, J. D. (2008), URL decay in MEDLINE—a 4-year follow-up study, Bioinformatics., 24(11), 1381 –1385, doi:10.1093/bioinformatics/btn127

