

# Umgang mit digitalen Forschungsdaten

## Initiativen in Deutschland

Heinz Pampel | Helmholtz-Gemeinschaft

Sitzung der AGEP der Humboldt-Universität zu Berlin, 25.06.2010

# Agenda

---

- Rahmen
- Akteure
- Aktivitäten
- Beispiel Geowissenschaften
- 10 Fragen
- 5 Prinzipien



# Kontext

---

- Bibliotheks- und Informationsmanagement (HdM)
- Helmholtz Open Access Projekt
  - Deutsches GeoForschungsZentrum GFZ
    - Abt. Bibliotheks- und Informationsdienste (LIS)
  - Alfred-Wegener-Institut für Polar- und Meeresforschung (AWI)
    - Computing and Data Centre
  - Arbeitspaket „Offener Zugang zu Forschungsdaten“

# Rahmen

---

- DFG: Sicherung guter wissenschaftlicher Praxis, 1998
- Allianz: Berliner Erklärung über offenen Zugang zu wissenschaftlichem Wissen, 2003
- Allianz: Schwerpunktinitiative „Digitale Information“, 2008
- DINI: Positionspapier Forschungsdaten, 2009
- DFG: Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten, 2009
- DFG: Leitfaden für Antragsteller, 2010

# Rahmen

---

- Deutsche Forschungsgemeinschaft (DFG), 1998
  - „Primärdaten als Grundlagen für Veröffentlichungen sollen auf haltbaren und gesicherten Trägern in der Institution, wo sie entstanden sind, für zehn Jahre aufbewahrt werden.“
  - „Auf die Aufzeichnungen später zurückgreifen zu können, ist schon aus Gründen der Arbeitsökonomie in einer Gruppe ein zwingendes Gebot. Noch wichtiger wird dies, wenn veröffentlichte Resultate von anderen angezweifelt werden.“

# Rahmen

---

- Berliner Erklärung, 2003
  - „Establishing open access as a worthwhile procedure ideally requires the active commitment of each and every individual producer of scientific knowledge and holder of cultural heritage. Open access contributions include original scientific research results, raw data and metadata, source materials, digital representations of pictorial and graphical materials and scholarly multimedia material.“

# Rahmen

---

- Schwerpunktinitiative „Digitale Information“, 2008
  - Die Aktivitäten der Allianz-Initiative sind auf drei Bereiche gerichtet:
    - Primärdaten-Policy
    - Zusammenarbeit zwischen Fachwissenschaftlern und Informationswissenschaftlern anzustoßen
    - Vernetzte Repositorien- und Archivstrukturen für Forschungsprimärdaten

# Rahmen

---

- DFG: Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten, 2009
  - „Die nun vorliegenden Empfehlungen richten sich sowohl an die Wissenschaftler als auch die Informationsspezialisten. Es soll damit das Bewusstsein für die Dringlichkeit dieses Problems geweckt werden. Darüber hinaus werden die Ziele benannt, die bei der Umsetzung berücksichtigt werden sollten und deren Einhaltung von grundlegender Bedeutung ist. Schließlich sollen die Wissenschaftler dazu angeregt werden, dieser Aufgabe auf dem Wege der Selbstverpflichtung nachzukommen und für eine sachgemäße Umsetzung im Rahmen der gegebenen Strukturen in eigener Initiative Sorge zutragen.“



# Rahmen

---

- DFG: Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten, 2009

- 1. Forschungsprimärdaten sind Daten, die im Verlauf von Quellenforschungen, Experimenten, Messungen, Erhebungen oder Umfragen entstanden sind. Sie stellen die Grundlagen für die wissenschaftlichen Publikationen dar.
- 2. Es ist ein fachspezifisches Organisationskonzept festzulegen, das die nachhaltige Speicherung der Daten regelt.
- 3. Die Speicherung der Forschungsdaten erfolgt im Rahmen definierter Standards
- 4. Die Daten sind persönlich gekennzeichnet und werden unter dem Namen des Forschers abgelegt.
- 5. Jeder Wissenschaftler stellt seine Forschungsprimärdaten nach Möglichkeit überregional und frei zur Verfügung.
- 6. Die Daten werden durch Metadaten beschrieben.
- 7. Die Daten sind qualitätsgeprüft.

# Rahmen

---

- DINI: Positionspapier Forschungsdaten, 2009
  - „Dieses Positionspapier beleuchtet – aus interdisziplinärer Sicht – fünf Kernthemen: Disziplinarität, Organisation, Zugang, Qualität und Technik, welche bei der Annäherung an das Themenfeld Forschungsdaten von grundlegender Bedeutung sind.“

# Rahmen

---

- DFG: Leitfaden für Antragsteller, 2010
  - „Wenn aus Projektmitteln systematisch (Mess-)Daten erhoben werden, die für die Nachnutzung geeignet sind, legen Sie bitte dar, welche Maßnahmen ergriffen wurden bzw. während der Laufzeit des Projektes getroffen werden, um die Daten nachhaltig zu sichern und ggf. für eine erneute Nutzung bereit zu stellen. Bitte berücksichtigen Sie dabei auch – sofern vorhanden – die in Ihrer Fachdisziplin existierenden Standards und die Angebote bestehender Datenrepositorien.“

# Akteure (Auswahl)

---

- Allianz AG Forschungsdaten
- DataCite (entstanden aus dem STD-DOI-Projekt)
- DINI AG Elektronisches Publizieren
- DINI AG Virtuelle Forschungsumgebungen
- Helmholtz Open Access Projekt
- ICSU World Data System (WDC-Climate, WDC-RSAT, WDC-MARE)
- KII AG Forschungsdaten
- Leibnitz AK Forschungsdaten

# Aktivitäten (Unvollständig)

---

- BW-eLabs und LiLa (u.a. U Stuttgart)
- D-GRID Initiative
- DataCite (u.a. TIB, ZB Med, ZBW)
- Earth System Science Data (u.a. AWI)
- Economics und NEEO – Networked of European Economists Online (u.a. ZBW)
- ICFA Study Group on Data Preservation and Long Term Analysis in HEP (u.a. DESY)
- KoLaWiss (U Göttingen)
- Konzeptstudie "Vernetzte Primärdaten-Infrastruktur für den Wissenschaftler-Arbeitsplatz in der Chemie,, (TIB)
- PANGAEA (AWI und MARUM)
- PlanktonNet (u.a. AWI)
- Scientific Drilling Database - SDDDB (GFZ)
- Publication and Citation of Scientific Primary Data - STD-DOI (WDCs und TIB)
- WDC-Climate (MPI für Meteorologie und Deutsches Klimarechenzentrum)
- WDC-RSAT (Deutsches Zentrum für Luft- und Raumfahrt)
- WDC-MARE (AWI und MARUM)
- Wibaklidama (FH Potsdam)
- World Atlas of Language Structures (MPDL)
- ...

# Beispiel Geowissenschaften

---

- Geokommission: Strategieschrift „Dynamische Erde“, 2010
- „Um die Erde nachhaltig nutzen zu können, müssen relevante Geodaten verfügbar sein. Diese Informationen sind Grundlage für regionale und globale Beobachtungs-, Vorhersage- und Entscheidungssysteme. Für das Management dieser Daten wurden Weltdatenbanken eingerichtet. Zudem gibt es auf regionaler und lokaler Ebene weitere Datenbanken für verschiedene Fachinformationen. Der standardisierte Zugriff auf diese Daten ermöglicht eine weltweite Verbreitung.“

# Beispiel Geowissenschaften

---

- Geokommission: Strategieschrift „Dynamische Erde“, 2010
- „Satellitenmessungen, Großrechner und hoch auflösende Analysemethoden produzieren täglich ungeheure Datenmengen. Diese Daten müssen besser zugänglich sein, damit sie zur Erforschung des Systems Erde beitragen können. Die Geowissenschaften müssen daher auf nationaler und internationaler Ebene ein einheitliches Datenmanagement aufbauen.“

# Beispiel Geowissenschaften

---

- Geokommission: Strategieschrift „Dynamische Erde“, 2010
- „Durch die eingeschränkte Nutzung ergeben sich Probleme: Weil wissenschaftliche Daten nicht auffindbar sind oder nicht zur Verfügung stehen, wird die interdisziplinäre und überregionale Forschung behindert. Die Forschung zum globalen Wandel benötigt zum Beispiel gute und ausführlich dokumentierte Daten. Existierende Datenbestände bleiben ungenutzt, da sie oftmals nur einem kleinen Kreis von Wissenschaftlern bekannt und zugänglich sind. Außerdem ist es häufig schwierig, veröffentlichte Ergebnisse zu verifizieren.“



# Beispiel Geowissenschaften

---

- Geokommission: Strategieschrift „Dynamische Erde“, 2010
- „Bei der Archivierung und Publikation großer Datenmengen gibt es folgende Probleme:
  - Die meisten Einrichtungen sind mit der Langzeitarchivierung der Daten überfordert.
  - Die Aufbereitung und Dokumentation von Daten ist zeit- und kostenintensiv.
  - Es fehlt ein Anreiz, Daten zu veröffentlichen und für die Veröffentlichung aufzubereiten.
  - Es fehlen anerkannte elektronische Medien, in denen Daten parallel zu deren Interpretation in traditionellen wissenschaftlichen Medien veröffentlicht werden können.“


# Beispiel Geowissenschaften

---

- Geokommission: Strategieschrift „Dynamische Erde“, 2010
- „Diese Defizite können durch folgende Maßnahmen behoben werden, die vor allem im Datenmanagement liegen:
  - Archivierung in Langzeit-Datenzentren („Datenbibliotheken“).
  - Metadaten und Ontologien zur Datenbeschreibung müssen entwickelt werden.
  - Nutzerfreundliche Informationsportale („one stop shop“) für den allgemeinen Zugang zu geowissenschaftlichen Daten.
  - Datenmanagement muss obligatorischer Bestandteil von Forschung, Lehre und Forschungsförderung sein.
  - Es muss ein System zur Publikation und Zitierbarkeit wissenschaftlicher Daten eingerichtet werden, inklusive Nutzung neuer Medien.
  - Effiziente Kostenmodelle für die Langzeitarchivierung und Publikation von Daten müssen formuliert werden.

# Beispiel Geowissenschaften

- PANGAEA

 **PANGAEA®**  
Publishing Network for Geoscientific & Environmental Data

You are not logged in (LOG IN)

Always quote citation when using data! [RIS](#) [BIBTeX](#)

### Data Description

**Citation:** Brückner, S; Mackensen, A (2008): Stable carbon isotope composition of benthic foraminifera from sediments of the Skagerrak, North Sea. doi:10.1594/PANGAEA.676719.  
**Supplement to:** Brückner, Sylvia; Mackensen, Andreas (2008): Organic matter rain rates, oxygen availability, and vital effects from benthic foraminiferal  $\delta^{13}C$  in the historic Skagerrak, North Sea. *Marine Micropaleontology*, 66(3-4), 192-207, doi:10.1016/j.marmicro.2007.09.002

**Reference(s):** Brückner, Sylvia (2008): Climatic and hydrographic variability in the late Holocene Skagerrak as deduced from benthic foraminiferal proxies (Klimatische und hydrographische Variabilität im holozänen Sagerrak, abgeleitet aus benthischen Foraminiferen). *Reports on Polar and Marine Research, Alfred Wegener Institute for Polar and Marine Research, Bremerhaven*, 572, 139 pp, hdi:10013/epic.28879.d001

**Abstract:** The sediment cores 225514 and 225510 were recovered from 420 and 285 m water depth, respectively. They were investigated for their benthic foraminiferal  $\delta^{13}C$  during the last 500 years. Both cores were recovered from the southern flank of the Skagerrak. The  $\delta^{13}C$  values of *Uvigerina mediterranea* and other shallow infaunal species in both cores indicate that organic matter rain rates to the seafloor varied around a mean value until approximately AD 1950 after which they increased. This increase might result from changes in the North Atlantic Current System and a co-occurring persistently high North Atlantic Oscillation index state in the 1980s to 1990s, rather than from anthropogenic eutrophication. Using  $\delta^{13}C$  mean values of multiple species, we reconstruct  $\delta^{13}C$  gradients of dissolved inorganic carbon (DIC) within pore waters for the time periods AD 1500 to 1950 and AD 1950 to 2000. The calculated  $\delta^{13}C$  ranges, interpreted as indicating total organic matter remineralization due to respiration, are generally bigger in Core 225514 than in Core 225510. Since mean  $\delta^{13}C$  values of *U. mediterranea* suggest that organic matter rain rates were similar at both locations, differences in total organic matter remineralization are attributed to differing oxygen availability. However, oxygen concentrations in the overlying bottom water masses are not likely to have differed significantly. Thus, we suggest that organic matter remineralization was controlled by oxygen availability within the sediments, reflecting strong differences in sedimentation rates at the two investigated core sites. Based on the assumptions that tests of benthic foraminiferal species inhabiting the same microhabitat depth should show equal  $\delta^{13}C$  values unless they are affected by vital effects and that *Globobulimina turgida* records pore water  $\delta^{13}C$ , we estimate microhabitat-corrected vital effects for several species with respect to *G. turgida*:  $>0.7$  per mil for *Cassidulina laevigata*,  $>1.3$  per mil for *Hyalinea balthica*, and  $>0.7$  per mil for *Melonis barleeanus*. *Melonis zaandami* seems to closely record pore water  $\delta^{13}C$ .

**Project(s):** [Paleoenvironmental Reconstructions from Marine Sediments @ AWI \(AWI\\_Paleo\)](#)

**Coverage:** West: 8.7091 \* East: 9.6214 \* South: 57.8406 \* North: 58.0422

**Event(s):** [Alkor\\_159\\_225510](#) \* [Latitude: 58.0422](#) \* [Longitude: 9.6214](#) \* [Elevation: -285.0 m](#) \* [Recovery: 5.44 m](#) \* [Location: Skagerrak](#) \* [Campaign: AL19xx](#) \* [Basis: Alkor](#) \* [Device: Gravity corer](#)  
[Alkor\\_159\\_225514](#) \* [Latitude: 57.8406](#) \* [Longitude: 8.7091](#) \* [Elevation: -420.0 m](#) \* [Recovery: 4.05 m](#) \* [Location: Skagerrak](#) \* [Campaign: AL19xx](#) \* [Basis: Alkor](#) \* [Device: Gravity corer](#)

**Comment:** Project: Integrated Baltic Sea Environmental Study (IBSEN), funded by German Ministry of Education and Research (BMBF)

**Size:** 5 datasets

### Download Data

Download ZIP file containing all datasets as tab-delimited text (use the following character encoding: [ISO-8859-1: ISO Western \(PANGAEA default\)](#))

### Datasets listed in this Collection

1. Brückner, S; Mackensen, A (2008): (Figure 3a) Stable carbon isotope ratios of benthic foraminifera from sediment core Alkor\_159\_225514. doi:10.1594/PANGAEA.676718
2. Brückner, S; Mackensen, A (2008): (Figure 3b) Stable carbon isotope ratios of benthic foraminifera from sediment core Alkor\_159\_225510. doi:10.1594/PANGAEA.676717
3. Brückner, S; Mackensen, A (2008): (Table 1) Age determination of sediment core Alkor\_159\_225510. doi:10.1594/PANGAEA.676716
4. Brückner, S; Mackensen, A (2008): (Table 3a) Paired  $\delta^{13}C$  differences between *Globobulimina turgida* and *Melonis barleeanus* of sediment core Alkor\_159\_225514. doi:10.1594/PANGAEA.716157
5. Brückner, S; Mackensen, A (2008): (Table 3b) Paired  $\delta^{13}C$  differences between various benthic foraminifera of sediment core Alkor\_159\_225510. doi:10.1594/PANGAEA.716158

Contact:

# Beispiel Geowissenschaften

- ScienceDirect - PANGAEA

The screenshot displays the ScienceDirect interface. At the top, it indicates 'You have Guest access to ScienceDirect' with a 'Find out more...' link. The navigation bar includes 'Home', 'Browse', 'Search', 'My Settings', 'Alerts', and 'Help'. Below this is a search bar with fields for 'All fields', 'Author', 'Journal/book title', 'Volume', 'Issue', and 'Page', along with 'Clear' and 'Go' buttons. The main content area shows a search result for the paper 'Organic matter rain rates, oxygen availability, and vital effects from benthic foraminiferal  $\delta^{13}\text{C}$  in the historic Skagerrak, North Sea' by Sylvia Brückner and Andreas Mackensen. The abstract is visible, and there are options to 'Purchase PDF (743 K)', 'Export Citation', and 'E-mail Article'. A 'Related Articles' sidebar on the right lists several other papers. At the bottom right, there is a 'Supplementary Data' section with a 'View Record in Scopus' link.

# Beispiel Geowissenschaften

- Bibliotheksverbund - PANGAEA

GBV GVK

Suchen [und] [ALL] Alle Wörter sortiert nach Erscheinungsjahr

Skagerrak Brückner   Unscharfe Suche

Benutzerkennung: GFZ POTSDAM | Abmelden/Datenbankmenü

Suchgeschichte | Kurzliste | **Titelanzeige** Copyright © 2010 OCLC

Ihre Aktion **Stable oxygen isotope composition of benthic foraminifera from sediments of the Skagerrak, North Sea, supplementary data to: Brückner, Sylvia; Mackensen, Andreas (2006): Deep-water renewal in the Skagerrak during the last 1200 years triggered by the North Atlantic Oscillation: evidence from benthic foraminiferal  $\delta^{18}O$ . Holocene, 16(3), 331-340** | [Trefferanalyse](#) | [Speichern/Drucken](#) | [Merken](#)  
1 Treffer

PPN: 572685475  
Titel: [Stable oxygen isotope composition of benthic foraminifera from sediments of the Skagerrak, North Sea, supplementary data to: Brückner, Sylvia; Mackensen, Andreas \(2006\): Deep-water renewal in the Skagerrak during the last 1200 years triggered by the North Atlantic Oscillation: evidence from benthic foraminiferal  \$\delta^{18}O\$ . Holocene, 16\(3\), 331-340](#)  
Verfasser: Brückner, Sylvia ; Mackensen, Andreas  
Erschienen: 2006-01-11  
Umfang: 2 Datasets.  
Anmerkung: This dataset is supplement to doi:10.1191/0959683605h931rp

Inhalt: Benthic foraminiferal tests of a sediment core from southwestern Skagerrak (northeastern North Sea, 420 m water depth) were investigated for their ratio of stable oxygen isotopes.  
Technische Angaben: Format: application/zip

Elektron. Referenz: <http://dx.doi.org/10.1594/PANGAEA.676724>

[Über den Zitierlink können Sie diesen Titel als Lesezeichen ablegen oder weiterleiten]

Besitzende Bibliothek(en): Klicken Sie auf einen Bibliotheksnamen oder auf >

**Hannover, TIB/UB Hannover <89>**

Bitte beachten Sie, dass nicht immer alle Materialien tatsächlich für die Ausleihe zur Verfügung stehen. Weitere Informationen erhalten Sie über den Katalog der jeweiligen Bibliothek durch Klick auf OPC.  
subito Lieferbibliotheken sind in Rot dargestellt

# Beispiel Geowissenschaften

- Earth System Science Data (ESSD)

The screenshot shows the Earth System Science Data (ESSD) website interface. The header includes the journal title and a 'Contact' link. A navigation menu on the left lists options like 'Home', 'Online Library ESSD', and 'Alerts & RSS Feeds'. The main content area features a paper by G. König-Langlo and H. Gernandt, with an abstract and a 'Discussion Paper' link. The right sidebar contains search boxes and a 'Recent Papers' list.

Earth System Science Data  
The Data Publishing Journal

| Contact |

Home  
Online Library ESSD  
Online Library ESSDD

- ▣ Papers in Open Discussion
- ▣ Volumes and Issues
- ▣ Special Issues
- ▣ Library Search
- ▣ Title and Author Search

Alerts & RSS Feeds  
General Information  
Submission  
Review  
Production  
Subscription  
Comment on a Paper

ARCHIVED IN  
PORTICO

Earth Syst. Sci. Data Discuss., 1, 1-13, 2008  
www.earth-syst-sci-data-discuss.net/1/1/2008/  
© Author(s) 2008. This work is distributed under the Creative Commons Attribution 3.0 License.

▣ Volumes and Issues ▣ Contents of Issue 1

**Compilation of ozonesonde 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, Bussestraße 24, 27570 Bremerhaven, Germany

**Abstract.** On 22 May 1985 the first balloon-borne ozonesonde was successfully launched by the staff of Georg-Forster-Station (70°46' S, 11°41' E). The following weekly ozone soundings mark the beginning of the continuous investigation of Germany to study the vertical ozone distribution in the southern hemisphere.

In 1985 these ozone soundings have been the only record showing the change of vertical ozone distribution in the southern polar stratosphere in September and October. 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 neighbouring station Novolazarevskaya, just 2 km apart from Georg-Forster-Station. Till 1992, ozone soundings were taken without interruption. Afterwards, the ozone sounding program was moved to Neumayer-Station (70°39' S, 8°15' W) 750 km further west.

▣ Discussion Paper (PDF, 423 KB) ▣ Interactive Discussion (Closed, 7 Comments) ▣ Final Revised Paper (ESSD)

Copernicus Publications  
The Innovative Open Access Publisher

Search ESSD

Library Search ▶▶  
Author Search ▶▶

Recent Papers

**02 | ESSDD, 30 Mar 2010:**  
EPOCA/EUR-OCEANS data-mining compilation on the impacts of ocean acidification

**03 | ESSD, 18 Mar 2010:**  
The CARINA data synthesis project: introduction and overview

**04 | ESSD, 15 Mar 2010:**  
The Irminger Sea and the Iceland Sea time series measurements of sea water carbon and nutrient chemistry 1983–2008

**05 | ESSDD, 03 Mar 2010:**

# Beispiel Geowissenschaften

- Earth System Science Data (ESSD)



## Data coverage and parameter measured

Repository-Reference: doi:10.1594/PANGAEA.547983  
Coverage: East: 11.8300; South: -70.7700;  
Location Name: Georg-Forster-Station, Antarctica  
Date/Time Start: 1985-05-22T05:19:00  
Date/Time End: 1992-01-29T01:19:00

# Beispiel Geowissenschaften

- Google Scholar - ESSD

The screenshot shows a Google Scholar search interface. The search bar contains the text "Compilation of ozonesonde profiles from the Antarctic". Below the search bar, there are options for "Web-Suche" (selected) and "Suche Seiten auf Deutsch". The search results are displayed in a list format. The first two results are identical, showing a PDF document titled "Compilation of ozonesonde profiles from the Antarctic Georg-Forster-Station ..." from earth-syst-sci-data-discuss.net. The third result is an abstract for the same topic, also from earth-syst-sci-data-discuss.net. The fourth result is a document titled "426 ozonesonde profiles from Georg-Forster-Station" from dx.doi.org. The interface includes a "Scholar" logo, a "Suche" button, and a "Erweiterte Scholar-Suche" link.

Google scholar   [Erweiterte Scholar-Suche](#)  
[Scholar-Einstellungen](#)  
 Web-Suche  Suche Seiten auf Deutsch

Scholar   Ergebnisse 1 - 4 von 4. (0,09 Sek.)

Tipp: [Suchen nur nach Ergebnissen auf Deutsch](#). Sie können Ihre bevorzugten Spracheinstellungen in [Scholar-Einstellungen](#) angeben.

[PDF] [Compilation of ozonesonde profiles from the Antarctic Georg-Forster-Station ...](#) [earth-syst-sci-data-discuss.net](#) [PDF]  
G Bremerhaven - 2008 - earth-syst-sci-data-discuss.net  
Earth Syst. Sci. Data Discuss., 1, 1–13, 2008 www.earth-syst-sci-data-discuss.net/1/1/2008/ ©  
Author(s) 2008. This work is distributed under the Creative Commons Attribution 3.0  
License. ... Earth System Science Data Discussions is the access reviewed discussion ...  
[Ähnliche Artikel](#) - [HTML-Version](#)

[PDF] [Compilation of ozonesonde profiles from the Antarctic Georg-Forster-Station ...](#) [earth-syst-sci-data-discuss.net](#) [PDF]  
G König-Langlo, H Gernandt - Earth System ..., 2008 - earth-syst-sci-data-discuss.net  
Earth Syst. Sci. Data Discuss., 1, 1–13, 2008 www.earth-syst-sci-data-discuss.net/1/1/2008/ ©  
Author(s) 2008. This work is distributed under the Creative Commons Attribution 3.0  
License. ... Earth System Science Data Discussions is the access reviewed discussion ...  
[Ähnliche Artikel](#) - [HTML-Version](#) - [Alle 7 Versionen](#)

[Compilation of ozonesonde profiles from the Antarctic Georg-Forster-Station from ...](#)  
G König-Langlo, H Gernandt - Sci. Data Discuss, 2008 - earth-syst-sci-data-discuss.net  
Abstract. On 22 May 1985 the first balloon-borne **ozonesonde** was successfully launched by  
the staff of **Georg-Forster-Station** (70°46' S, 11°41' E). The following weekly ozone soundings  
mark the **beginning of the continuous investigation of Germany to study the vertical ozone ...**  
[Im Cache](#) - [Alle 3 Versionen](#)

[426 ozonesonde profiles from Georg-Forster-Station](#)  
G König-Langlo, H Gernandt - Alfred Wegener Institute for Polar and ..., 2008 - dx.doi.org  
... Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, doi:10.1594/PANGAEA.  
547983, Supplement to: König-Langlo, Gert; Gernandt, Hartwig (2008): **Compilation of  
ozonesonde profiles from the Antarctic Georg-Forster-Station from 1985 to 1992.** ...  
[Alle 2 Versionen](#)



# Beispiel Geowissenschaften

- Scientific Drilling Database (SDDDB)

The screenshot displays the Scientific Drilling Database (SDDDB) interface. The header includes the title 'Scientific Drilling Database' and the subtitle 'Data from Deep Earth Sampling and Monitoring'. A left sidebar contains navigation links such as Home, About SDDDB, News, Data Publications, Catalogue, Authors, Dataset, Research Programs, Sampling Gear, Analytical Parameters, Publication Process, and Admin. The main content area is titled 'Dataset Description' and provides the following information:

- Citation:** Demory, Francois; Oberhänsli, Hedi; Nowaczyk, Norbert; Gottschalk, Matthias; Wirth, Richard; Naumann, Rudolf; (2006): Hysteresis measurements (uncorrected and corrected for the paramagnetic influence) for VER98-1-1 at core depth of 580.6cm. *Scientific Drilling Database*. doi:10.1594/GFZ.SDDDB.1031
- DOI:** 10.1594/GFZ.SDDDB.1031
- Title:** Hysteresis measurements (uncorrected and corrected for the paramagnetic influence) for VER98-1-1 at core depth of 580.6cm.
- Abstract:** Hysteresis measurements of a hematite dominated diatomaceous layer with pseudosingle to multidomain behaviour of the hysteresis loop for VER98-1-1.
- Related Publications:** Francois Demory, Hedi Oberhänsli, Norbert R. Nowaczyk, Matthias Gottschalk, Richard Wirth and Rudolf Naumann, Detrital input and early diagenesis in sediments from Lake Baikal revealed by rock magnetism, *Global and Planetary Change*, Volume 46, Issues 1-4, doi:10.1016/j.gloplacha.2004.11.010
- Activities:** VER98-1-1
  - Latitude: 53.3933 °N
  - Longitude: 107.9227 °E
  - Elevation: -245 m above site datum
  - Date/Time: 1998-09-03 00:00:00 UTC
  - Program: Vereshchagin Expedition 98
  - Expedition: VER98-1
  - Platform: R/V Vereshchagin
  - Gear: Piston corer Meischner
- Datapoints:** 1512

Additional features include a 'Download Citation (EndNote)' button, a 'Show in Google Earth' button, and a right sidebar with 'Glossary' and 'Catalogue' links. Logos for GFZ Potsdam and ICDP are visible in the bottom left of the page.

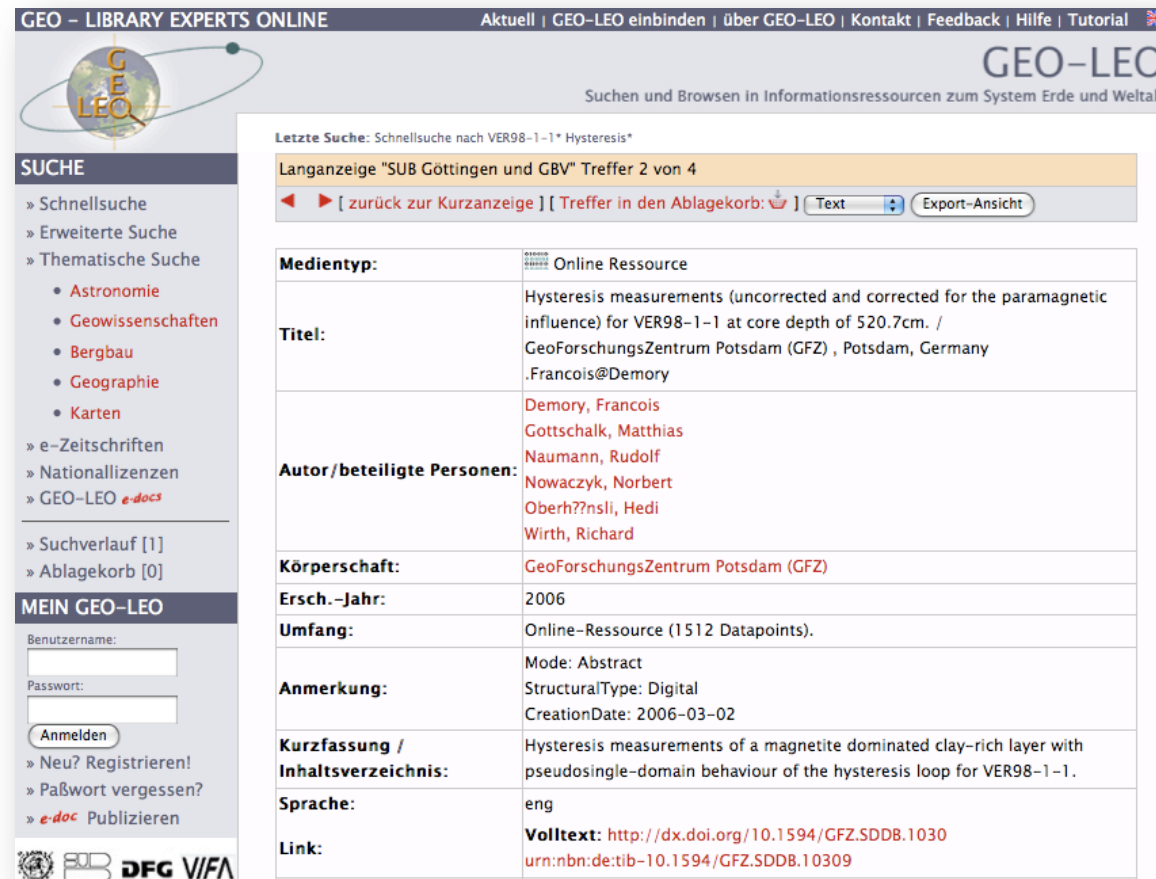
# Beispiel Geowissenschaften

- Bibliothekskatalog - SDDB

The screenshot displays the ALBERT library catalog interface. At the top, there is a navigation bar with options: Simple Search, Advanced Search, Journals A-Z, Mind List (0), Search History, and Settings. Below this is a search input field containing the query 'uid:sddb\_10.1594/GFZ.SDDB.1031' and a 'GO' button. The search results show '1 hit in 0.019 seconds'. On the left, a 'Refine your search' sidebar lists 'Collection' with 'All Sources (1)', 'Books (0)', 'Journals (0)', 'Articles (0)', 'Data (1)', and 'More (0)'. The main search results area shows a single entry: '1.  **Hysteresis measurements (uncorrected and corrected for the paramagnetic influence) for VER98-1-1 at core depth of 580.6cm.** (2006) [DATA SDDB](#)'. The entry is by Francois Demory, Matthias Gottschalk, and Rudolf Naumann, published by Deutsches GeoForschungsZentrum - GFZ. It includes links for 'Show Details', '10.1594/GFZ.SDDB.1031', 'S·F·X', and 'Permalink'. The description states: 'Hysteresis measurements of a hematite dominated diatomaceous layer with pseudosingle to multidomain behaviour of the hysteresis loop for VER98-1-1.' Other metadata includes: Publication Date: 2006-09-15, Language: en, Structural Type: Digital, Mode: Abstract, Resource Type: Dataset, Registration Agency: 10.1594, Issue Date: 2006-09-15, Issue Number: 2, Creation Date: 2006-03-02, Data Description: text/tab-separated-values, and Format: Datapoints.

# Beispiel Geowissenschaften

- GEO-LEO - SDDB



GEO – LIBRARY EXPERTS ONLINE    Aktuell | GEO-LEO einbinden | über GEO-LEO | Kontakt | Feedback | Hilfe | Tutorial

GEO-LEO  
Suchen und Browsen in Informationsressourcen zum System Erde und Weltall

Letzte Suche: Schnellsuche nach VER98-1-1\* Hysteresis\*

Langanzeige "SUB Göttingen und GBV" Treffer 2 von 4

◀ ▶ [ zurück zur Kurzanzeige ] [ Treffer in den Ablagekorb: 📁 ] [ Text ] [ Export-Ansicht ]

<b>Medientyp:</b>	Online Ressource
<b>Titel:</b>	Hysteresis measurements (uncorrected and corrected for the paramagnetic influence) for VER98-1-1 at core depth of 520.7cm. / GeoForschungsZentrum Potsdam (GFZ) , Potsdam, Germany .Francois@Demory
<b>Autor/beteiligte Personen:</b>	Demory, Francois Gottschalk, Matthias Naumann, Rudolf Nowaczyk, Norbert Oberh??nsl, Hedi Wirth, Richard
<b>Körperschaft:</b>	GeoForschungsZentrum Potsdam (GFZ)
<b>Ersch.-Jahr:</b>	2006
<b>Umfang:</b>	Online-Ressource (1512 Datapoints).
<b>Anmerkung:</b>	Mode: Abstract StructuralType: Digital CreationDate: 2006-03-02
<b>Kurzfassung / Inhaltsverzeichnis:</b>	Hysteresis measurements of a magnetite dominated clay-rich layer with pseudosingle-domain behaviour of the hysteresis loop for VER98-1-1.
<b>Sprache:</b>	eng
<b>Link:</b>	<b>Volltext:</b> <a href="http://dx.doi.org/10.1594/GFZ.SDDB.1030">http://dx.doi.org/10.1594/GFZ.SDDB.1030</a> <a href="urn:nbn:de:tib-10.1594/GFZ.SDDB.10309">urn:nbn:de:tib-10.1594/GFZ.SDDB.10309</a>

DFG VIFA

# 10 Fragen

PURDUE  
UNIVERSITY

**Libraries** Access. Knowledge. Success.

## INTRODUCTION

Librarians at Purdue University are beginning to identify the scientific datasets that are being generated by our faculty and researchers as information assets to be collected, preserved, and made accessible as a function of the library's collection development. These librarians are subject-area specialists, and many have advanced degrees in their respective disciplines in addition to a degree in library science. They have all been trained in collection management, however, much of this training was related to traditional formats such as monographs and serials and not datasets. In our experience, one of the most effective tactics for eliciting datasets for the collection is a simple librarian-researcher interview. In this poster, we share a set of ten questions that a librarian can use as a starting point for such a "data interview". It is not a comprehensive strategy but instead a practical tool to draw out information that needs to be considered in order to evaluate the suitability of a dataset for the collection and the requirements for the infrastructure and services that will be needed for data curation.

### #1 What is the story of the data?

Begin the interview with an open-ended question that allows the researcher to talk freely about his or her research, scientific workflow, and community of practice. This lends some insight into the value of the dataset and how it may fit into your collection and be used, and it provides the context for understanding how and why the dataset was created and how it was processed and analyzed.

### #2 What form and format are the data in?

What computing environments (e.g., software) are required to use the data? If the data are in proprietary structures, you may consider reformatting them into agnostic formats or ones that can be more easily re-versioned. Is there any existing metadata, either external to the data or description that could be extracted from it? Ideally the data could be described to be discoverable by researchers from another discipline.

### #3 What is the expected lifespan of the dataset?

In many cases, there are distinctions in the utility of a dataset as it begins in a raw state and then is analyzed and processed into new forms and versions as a result of different steps in the research workflow. Different entities may have custody of the data and use it for different purposes at different times, affecting its provenance. Funding agencies may require that data be archived for a prescribed period of time or you may forecast its future value and the amount of time it should be retained. The data may be described and archived for effective preservation to ensure its accessibility and integrity over time.

### #4 How could the data be used, reused, and repurposed?

This is a primary selection criterion that also impacts how the data are accessed and what policies may be needed to govern its use. As data are archived and shared, new and unintended uses for the data may increase its value. For example, a research dataset may be repurposed as a learning object.

### #5 How large is the dataset, and what is its rate of growth?

It is important to quantify the size of the data for storage and network provisioning if you intend to ingest it into your repository. What is its physical

(bits) and logical (records) scale? Is the dataset static or dynamic? Ask for a sample of the data to examine.

### #6 Who are the potential audiences for the data?

Information regarding potential users of the data and the users' needs is paramount. Along with potential uses for the data, this is another primary selection criterion. In some cases, the data may need to be embargoed or restricted to a limited group of users who are granted permission to access it.

### #7 Who owns the data?

Establishing and maintaining the intellectual property represented by the data should be discussed at the earliest opportunity, and any conflicts should be resolved up-front. Many organizations have a submission policy that asks the contributor to verify that they own the data and have the right to submit it.

### #8 Does the dataset include any sensitive information?

All data should be reviewed for information that violates confidentiality, such as identification information on human subjects. Data curation activities should be informed by institutional review board requirements.

### #9 What publications or discoveries have resulted from the data?

The researchers may have a bias regarding the importance of their data. The purpose of this question is to establish an objective metric for determining the value of the data for the collection. Different metrics may be more appropriate in determining the selection criteria for different kinds of data and data collections.

### #10 How should the data be made accessible?

There is value in making data accessible using a conventional web-based user interface, but machine-to-machine interfaces should also be evaluated. These methods of access will be informed by the answers to the previous questions, and this question can be asked in an open-ended manner to fill in any gaps remaining at the conclusion of the interview.

## SUMMARY

Although building robust collections of datasets present several complexities and challenges to resolve, the process of looking at scientific datasets as information assets and exploring what is needed to develop and manage data collections is similar to the traditional collection development practices that have been successfully employed by librarians for decades. We offer these ten "data interview" questions as a springboard for librarians to explore data curation in greater depth and specialization.

Michael Witt ([mwitt@purdue.edu](mailto:mwitt@purdue.edu))  
Assistant Professor of Library Science

Jake Carlson ([jrcarlso@purdue.edu](mailto:jrcarlso@purdue.edu))  
Data Research Scientist

Purdue University Libraries  
Distributed Data Curation Center  
<http://d2c2.lib.purdue.edu>



## "Conducting a Data Interview"

Michael Witt & Jake Carlson, Purdue University Libraries, West Lafayette, Indiana, USA

## INTRODUCTION

Librarians at Purdue University are beginning to identify the scientific datasets that are being generated by our faculty and researchers as information assets to be collected, preserved, and made accessible as a function of the library's collection development. These librarians are subject-area specialists, and many have advanced degrees in their respective disciplines in addition to a degree in library science. They have all been trained in collection management; however, much of this training was related to traditional formats such as monographs and serials and not datasets. In our experience, one of the most effective tactics for eliciting datasets for the collection is a simple librarian-researcher interview. In this poster, we share a set of ten questions that a librarian can use as a starting point for such a "data interview". It is not a comprehensive strategy but instead a practical tool to draw out information that needs to be considered in order to evaluate the suitability of a dataset for the collection and the requirements for the infrastructure and services that will be needed for data curation.

### #1 What is the story of the data?

Begin the interview with an open-ended question that allows the researcher to talk freely about his or her research, scientific workflow, and community of practice. This lends some insight into the value of the dataset and how it may fit into your collection and be used, and it provides the context for understanding how and why the dataset was created and how it was processed and analyzed.

### #2 What form and format are the data in?

What computing environments (e.g., software) are required to use the data? If the data are in proprietary structures, you may consider reformatting them into agnostic formats or ones that can be more easily re-versioned. Is there any existing metadata, either external to the data or description that could be extracted from it? Ideally the data could be described to be discoverable by researchers from another discipline.

### #3 What is the expected lifespan of the dataset?

In many cases, there are distinctions in the utility of a dataset as it begins in a raw state and then is analyzed and processed into new forms and versions as a result of different steps in the research workflow. Different entities may have custody of the data and use it for different purposes at different times, affecting its provenance. Funding agencies may require that data be archived for a prescribed period of time or you may forecast its future value and the amount of time it should be retained. The data may be described and archived for effective preservation to ensure its accessibility and integrity over time.

### #4 How could the data be used, reused, and repurposed?

This is a primary selection criterion that also impacts how the data are accessed and what policies may be needed to govern its use. As data are archived and shared, new and unintended uses for the data may increase its value. For example, a research dataset may be repurposed as a learning object.

### #5 How large is the dataset, and what is its rate of growth?

It is important to quantify the size of the data for storage and network provisioning if you intend to ingest it into your repository. What is its physical

(bits) and logical (records) scale? Is the dataset static or dynamic? Ask for a sample of the data to examine.

### #6 Who are the potential audiences for the data?

Information regarding potential users of the data and the users' needs is paramount. Along with potential uses for the data, this is another primary selection criterion. In some cases, the data may need to be embargoed or restricted to a limited group of users who are granted permission to access it.

### #7 Who owns the data?

Establishing and maintaining the intellectual property represented by the data should be discussed at the earliest opportunity, and any conflicts should be resolved up-front. Many organizations have a submission policy that asks the contributor to verify that they own the data and have the right to submit it.

### #8 Does the dataset include any sensitive information?

All data should be reviewed for information that violates confidentiality, such as identification information on human subjects. Data curation activities should be informed by institutional review board requirements.

### #9 What publications or discoveries have resulted from the data?

The researchers may have a bias regarding the importance of their data. The purpose of this question is to establish an objective metric for determining the value of the data for the collection. Different metrics may be more appropriate in determining the selection criteria for different kinds of data and data collections.

### #10 How should the data be made accessible?

There is value in making data accessible using a conventional web-based user interface, but machine-to-machine interfaces should also be evaluated. These methods of access will be informed by the answers to the previous questions, and this question can be asked in an open-ended manner to fill in any gaps remaining at the conclusion of the interview.

## SUMMARY

Although building robust collections of datasets present several complexities and challenges to resolve, the process of looking at scientific datasets as information assets and exploring what is needed to develop and manage data collections is similar to the traditional collection development practices that have been successfully employed by librarians for decades. We offer these ten "data interview" questions as a springboard for librarians to explore data curation in greater depth and specialization.

Michael Witt ([mwitt@purdue.edu](mailto:mwitt@purdue.edu))  
Assistant Professor of Library Science

Jake Carlson ([jrcarloso@purdue.edu](mailto:jrcarloso@purdue.edu))  
Data Research Scientist

Purdue University Libraries  
Distributed Data Curation Center  
<http://d2c2.lib.purdue.edu>



# "Conducting a Data Interview"

Michael Witt & Jake Carlson, Purdue University Libraries, West Lafayette, Indiana, USA

# 5 Prinzipien

---

- RIN: Stewardship of digital research data, 2007

- **Principle 1: Roles and Responsibilities**

“The roles and responsibilities of researchers, research institutions and funders should be defined as clearly as possible, and they should collaboratively establish a framework of codes of practice to ensure that creators and users of research data are aware of and fulfil their responsibilities in accordance with these principles.”

# 5 Prinzipien

---

- RIN: Stewardship of digital research data, 2007

- **Principle 2: Standards and Quality Assurance**

“Digital research data should be created and collected in accordance with applicable international standards, and the processes for selecting those to be made available to others should include proper quality assurance.”

# 5 Prinzipien

---

- RIN: Stewardship of digital research data, 2007

- **Principle 3: Access, Usage and Credit**

“Digital research data should be easy to find, and access should be provided in an environment which maximises ease of use; provides credit for and protects the rights of those who have gathered or created data; and protects the rights of those who have legitimate interests in how data are made accessible and used.”



# 5 Prinzipien

---

- RIN: Stewardship of digital research data, 2007

- **Principle 4: Benefits and Cost-Effectiveness**

“The models and mechanisms for managing and providing access to digital research data must be both efficient and cost-effective in the use of public and other funds.”

# 5 Prinzipien

---

- RIN: Stewardship of digital research data, 2007

- **Principle 5: Preservation and Sustainability**

“Digital research data of long term value arising from current and future research should be preserved and remain accessible for current and future generations.”

# Danke

---

Quellen und Literaturhinweise:

<http://www.connotea.org/user/pampel/tag/hu2010>

Heinz Pampel

[pampel@gfz-potsdam.de](mailto:pampel@gfz-potsdam.de)

<http://oa.helmholtz.de>

# Backup

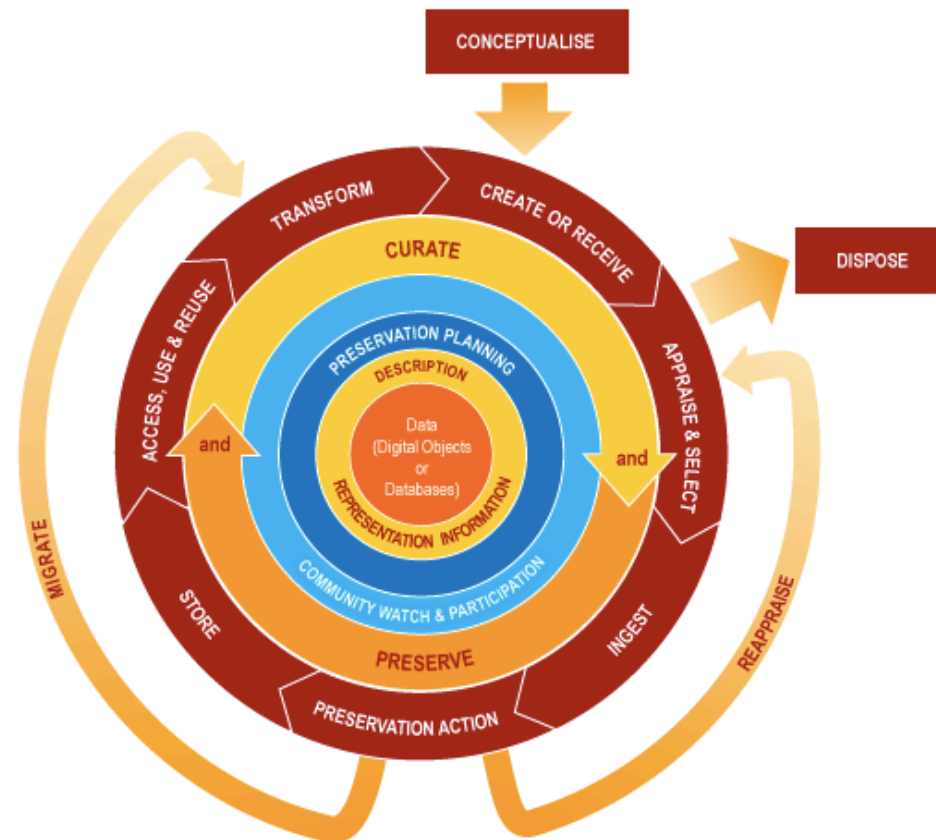
---

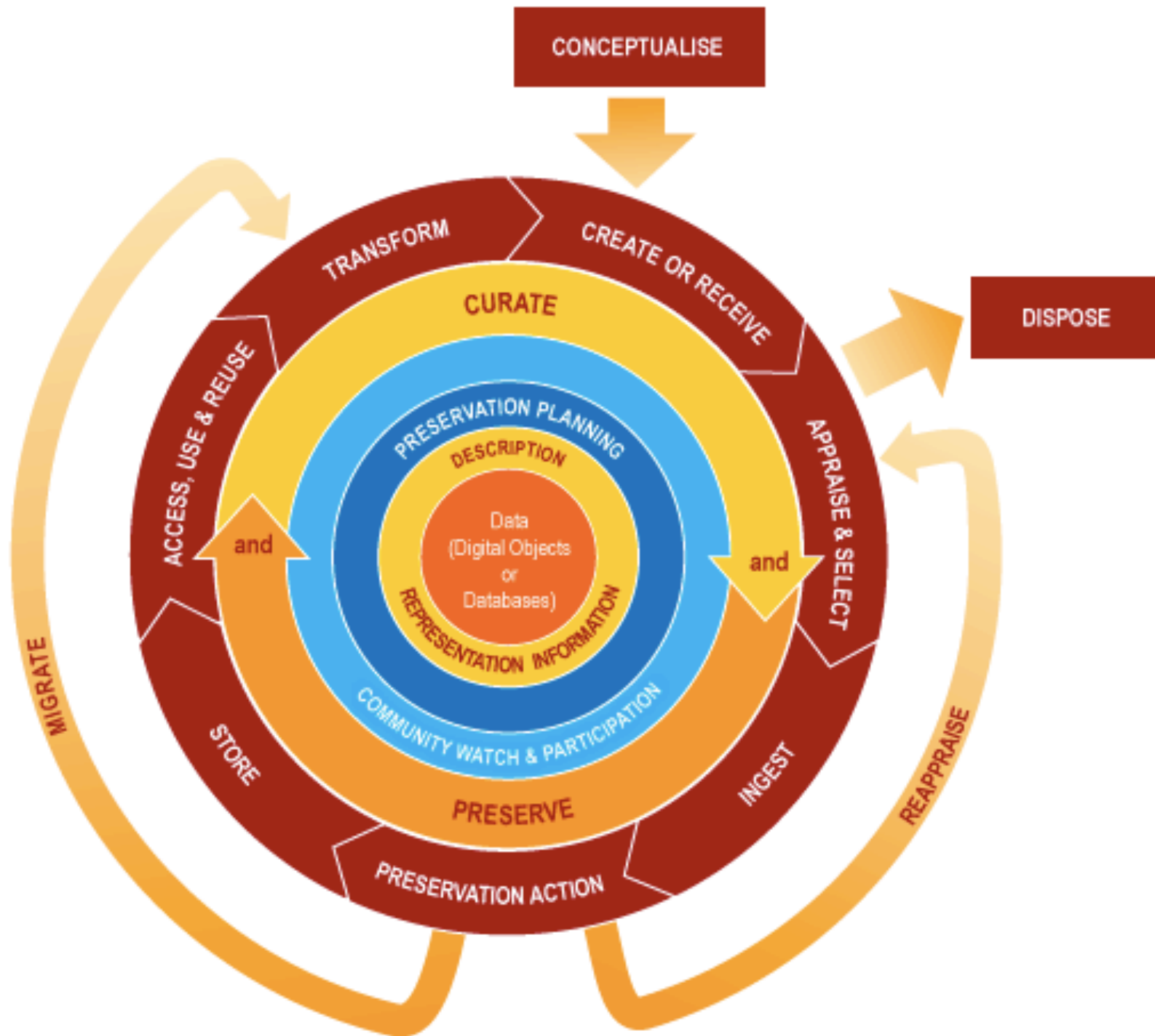
# NSB: Data Collections, 2005

---

- **Research Collections:** „Authors are individual investigators and investigator teams. Research collections are usually maintained to serve immediate group participants only for the life of a project, and are typically subjected to limited processing or curation. Data may not conform to any data standards.“
- **Resource Collections:** „Resource collections are authored by a community of investigators, often within a domain of science or engineering, and are often developed with communitylevel standards. Budgets are often intermediate in size. Lifetime is between the mid- and long-term.“
- **Reference Collections:** „Reference collections are authored by and serve large segments of the science and engineering community and conform to robust, well-established and comprehensive standards, which often lead to a universal standard. Budgets are large and are often derived from diverse sources with a view to indefinite support“

# DCC: Curation Lifecycle Model





# Treloar und Harboe-Ree: Data Curation Continuum, 2008

