

A blurred, 3D-rendered logo of the Deutsche Forschungsgemeinschaft (DFG) in white, set against a blue background with light rays.

Forschungsdaten und Fördermaßnahmen: Zielsetzung und Ansätze der DFG

Das Programm

- Einführung
- Wissenschaftlerinnen und Wissenschaftler und ihre Infrastrukturen
- Maßnahmen der DFG und anderer Wissenschaftsorganisationen
- Ein Fazit

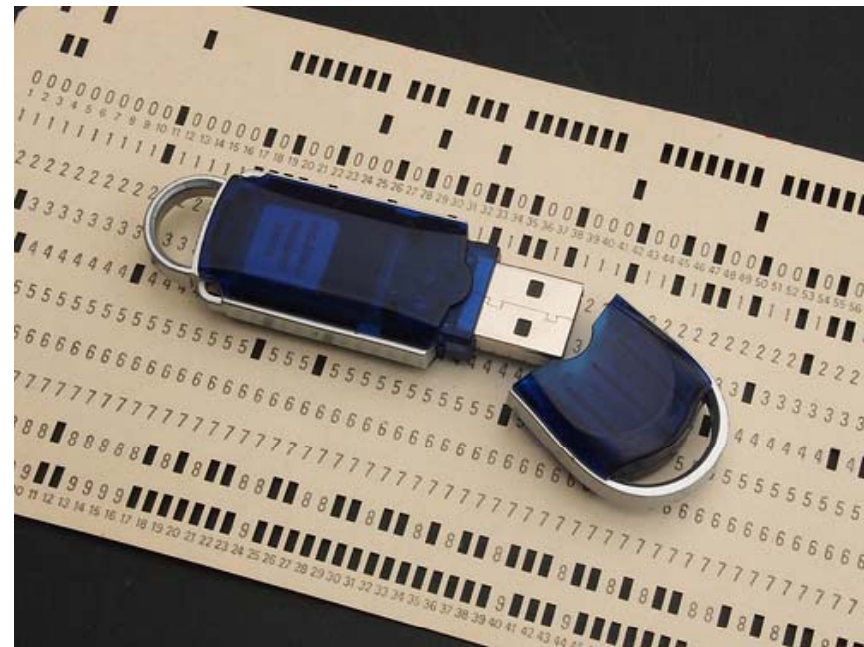


Quelle: DFG

1. Um was geht es? Forschungsdaten ...

Der Versuch einer Definition:

... Unter Forschungsdaten sind [...] digitale und elektronisch speicherbare Daten zu verstehen, die im Zuge eines wissenschaftlichen Vorhabens z.B. durch Quellenforschungen, Experimente, Messungen, Erhebungen oder Befragungen entstehen. ...



Ian-S; <http://www.flickr.com/photos/ian-s/2152798588/>; cc: by-nc-nd

2. Um was geht es? Forschungsdaten ...

- ... sind die Grundlage wissenschaftlicher Erkenntnis.
- ... werden heute nur unzureichend genutzt.
- ... sind nur eingeschränkt zugänglich.
- ... stehen nicht langfristig zur Verfügung.

Die Vision der Wissenschaftsorganisationen

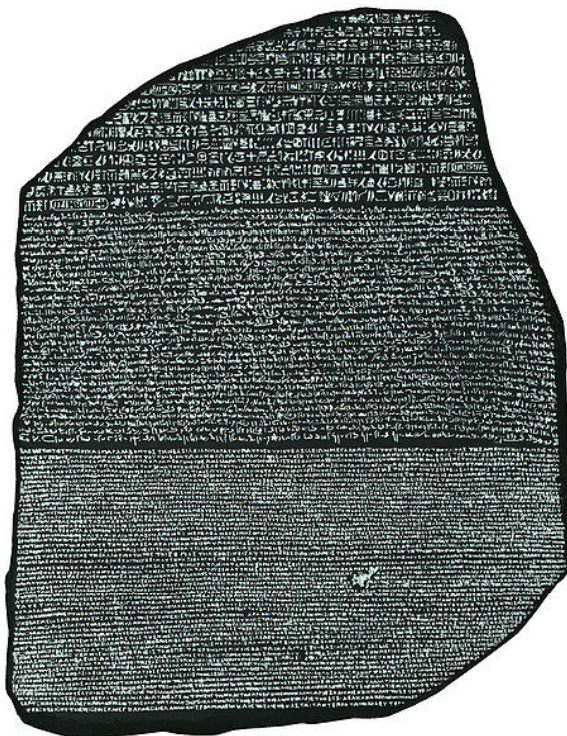
- Forschungsdaten sollen möglichst frei und überregional zugänglich und langfristig verfügbar sein.

Warum? Was soll bezweckt werden?

- Datenproduktion und Datennutzung soll generell optimiert werden.
- Ein offener Zugang ermöglicht neue Erkenntnisse, Forschungsansätze und Zusammenhänge, erschließt neue Einsatzbereiche, erlaubt die Betrachtung unter fremden Blickwinkel und ist Motor für Innovation.
- Forschungsprozess wird beschleunigt und nachhaltig verbessert.
- Erweiterung wissenschaftlichen Arbeitens: Experiment, Theorie, Modellierung und nun „data-driven science“.
- Ein offener Zugang zu Daten ist der Standard „Guter Wissenschaftlichen Praxis“.
- Der „dauerhafte Zugang zu qualitätsgesicherten Forschungsdaten“ ist von strategischer Bedeutung für die Wettbewerbsfähigkeit des Wissenschaftsstandorts.

Beispiele für die Lebenszeit von Information

Alter: 2.206 Jahre



http://upload.wikimedia.org/wikipedia/commons/c/ca/Rosetta_Stone_BW.jpeg

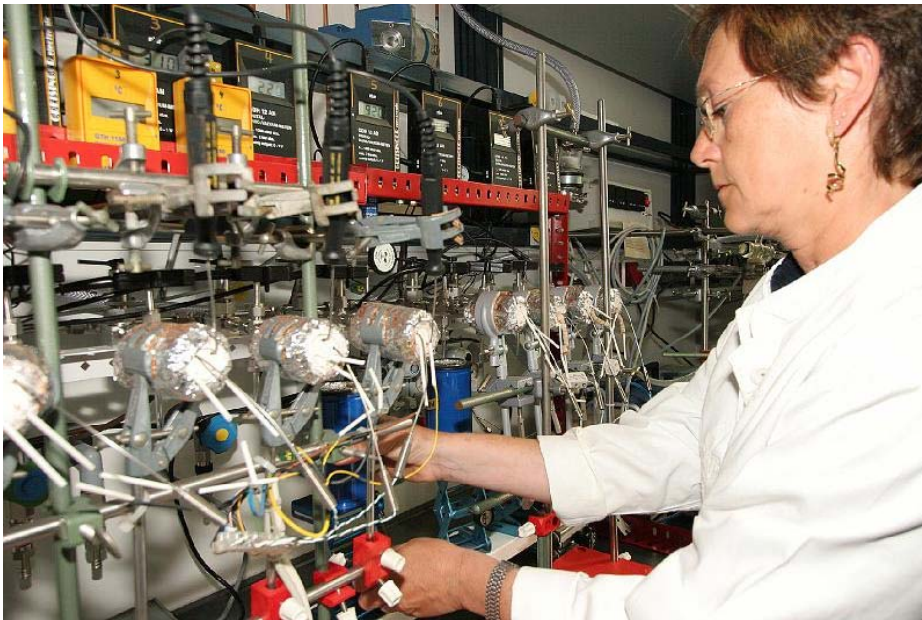
Alter: < 10 Jahre



Stinging Eyes; <http://www.flickr.com/photos/martinlatter/299981441> cc: by-sa

Das Programm, Teil 2

- Wissenschaftlerinnen und Wissenschaftler und ihre Infrastrukturen



Quelle: DFG

Wissenschaftliche Infrastruktur



Quelle: School of Chemistry, University Southampton



Quelle: Helmholtz Gemeinschaft



Quelle: DFG



Quelle: Helmholtz Gemeinschaft

Und auch Bibliotheken ...

4.000 Jahre Informationsmanagement ...



Der alte Lesesaal der British Library, London

http://en.wikipedia.org/wiki/File:British_Museum_Reading_Room_Panorama_Feb_2006.jpg cc: by-sa

... und Archive!



... Mit **Archivwürdigkeit** bezeichnet man im Archivwesen die Eigenschaft von zu übernehmenden Archivgut in ein historisches Endarchiv. Die Ermittlung der Archivwürdigkeit geschieht durch den Vorgang der archivischen Bewertung. Nicht archivwürdiges Archivgut wird in der Regel kassiert. **Unterlagen sind archivwürdig, wenn sie für Fragestellungen in der Zukunft auswertbar erscheinen. ...**

... sind Forschungsdaten „Kulturgut“?

"Archiv der Deutschen Abgeordneten", ein Kunstwerk von Christian Boltanski, im Untergeschoß des Reichstagesgebäudes
<http://www.flickr.com/photos/mr172/4451824642/>: cc: by-nc-sa

Zusammenspiel von hoch volumigen Archiven und Storageplattformen

11. Oracle Bibliotheken Summit, Weimar 2010

Stefan Luther

Leiter Kompetenzzentrum
Elektronische Dokumentenverwaltung
Nord Ost West Informationstechnik GmbH
Königsallee 71, 40194 Düsseldorf



*... Das Volumen in TB ist nicht kritisch,
Entscheidend ist die Anzahl der Files:*

1.856.786.299

*Die Akten der Deutschen Rentenversicherung
haben keinen zu definierenden Life Cycle ! ...*

Das Löschen ist ein Problem!

Und das Netz ...



Und noch viele offene Fragen

- **Wissenschaftliche Freiheit**
(GG §5). Auch die Freiheit, Daten nicht freizugeben?
- **Wem gehören die Daten?**
Förderorganisation, Institution/Einrichtung, (Haupt)-Antragsteller, Wissenschaftlicher Mitarbeiter, Verlag? Alles ist möglich.
- **Wie werden Daten zur Verfügbar gestellt?**
Was für Rechte werden für die Nachnutzung gewährleistet? Welche Lizenzen verwendet? Die „Offenheit“ von Daten ist nicht nur eine technische Frage.
- **Qualitätssicherung**
Alle Daten? Ist das überhaupt möglich? Wünschenswert? Wer entscheidet, welche Daten wo und wie lange gespeichert werden?

Und viele Bedenken ...

Es gibt „gute Gründe“ Daten nicht zu teilen, sprich „wegzugeben“:

- Darin stecken xxx Euro und yyy Jahre an Arbeit.
- Sie könnten so leicht missverstanden werden.
- Sie könnten fehlerhaft sein und jemand Fremdes könnte das feststellen.
- Irgendjemand könnte etwas Interessanteres darin finden.
- Irgendjemand könnte sie vor mir veröffentlichen.
- Es braucht viel zu viel Zeit, sie in einen neuen Kontext zu setzen und sie neu zu formatieren.
- Wie und wo soll ich meine Daten überhaupt einstellen?
- Ich habe ja überhaupt keine Kontrolle mehr über meine Daten.

... denn, das sind doch MEINE Daten!

... um so besser !?!

OPEN ACCESS Freely available online



Sharing Detailed Research Data Is Associated with Increased Citation Rate

Heather A. Piwowar*, Roger S. Day, Douglas B. Fridsma

Department of Biomedical Informatics, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, United States of America

Background. Sharing research data provides benefit to the general scientific community, but the benefit is less obvious for the investigator who makes his or her data available. **Principal Findings.** We examined the citation history of 85 cancer microarray clinical trial publications with respect to the availability of their data. The 48% of trials with publicly available microarray data received 85% of the aggregate citations. Publicly available data was significantly ($p = 0.006$) associated with a 69% increase in citations, independently of journal impact factor, date of publication, and author country of origin using linear regression. **Significance.** This correlation between publicly available data and increased literature impact may further motivate investigators to share their detailed research data.

Citation: Piwowar HA, Day RS, Fridsma DB (2007) Sharing Detailed Research Data Is Associated with Increased Citation Rate. PLoS ONE 2(3): e308. doi:10.1371/journal.pone.0000308

Infrastrukturen für Forschungsdaten: Beispiele 1

- AstroGrid-D: verteilte Forschungsdaten aus der Astronomie, sowie digitale Ressourcen

<http://www.gac-grid.de/>

The screenshot shows the AstroGrid-D website. At the top, there is a navigation bar with the AstroGrid-D logo, the text "German Astronomy Community Grid (GACG)", and a "D-GRID" logo. Below the navigation bar are tabs for "AstroGrid-D", "Project Info", "Documents", "Products", "Links", and "Intranet". A search bar is located on the right side of the navigation bar. The main content area features a heading "AstroGrid-D: enabling grid science in the German Astronomical community." followed by a description of the grid service. Below this, there is a section titled "Sponsored by" with the logo of the "Bundesministerium für Bildung und Forschung". The main content is organized into three sections: "Overview" with icons for "Grid Hardware", "Statistics", "Resource Maps", and "Job Timeline"; "Integrated Grid Services" with icons for "stellaris Information Server", "Virtual Organisations", and "Grid Way Meta Scheduler"; and "Scientific Applications" with icons for "Dynamo", "Nbody6++", "GEO600", "Clusterfinder", "Cactus", "Robotic Telescopes", and "ProC".

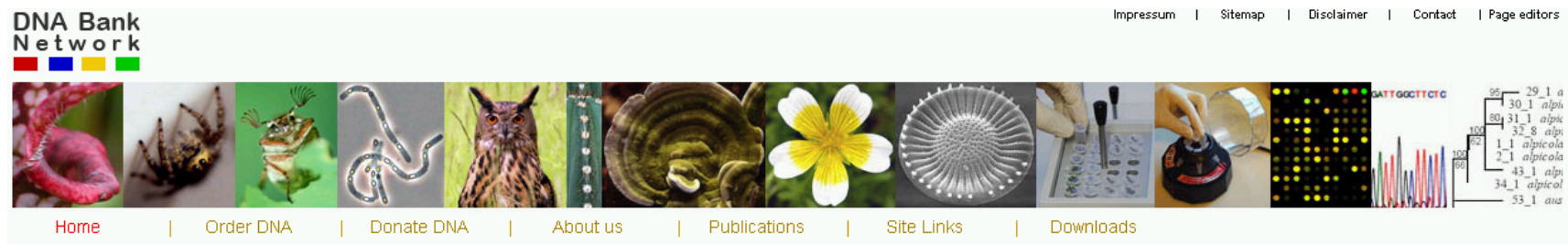
- ZPID/PsychData: Fachinformationszentrum mit Primärdaten der empirisch-psychologischen Forschung

<http://www.zpid.de/>

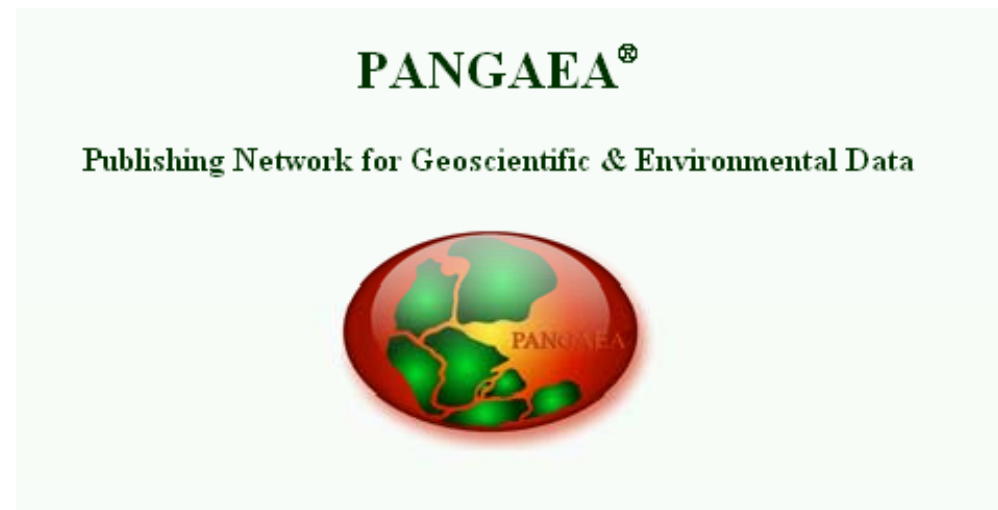


Infrastrukturen für Forschungsdaten: Beispiele 2

- DNA-Bank Netzwerk: Aufbau einer online Datenbank sowie physische Lagerung von DNA-Proben (<http://www.dnabank-network.org/>).



- PANGAEA: georeferenzierte Primärdaten aus der Erdsystem- und Umweltforschung (<http://www.pangaea.de/>).

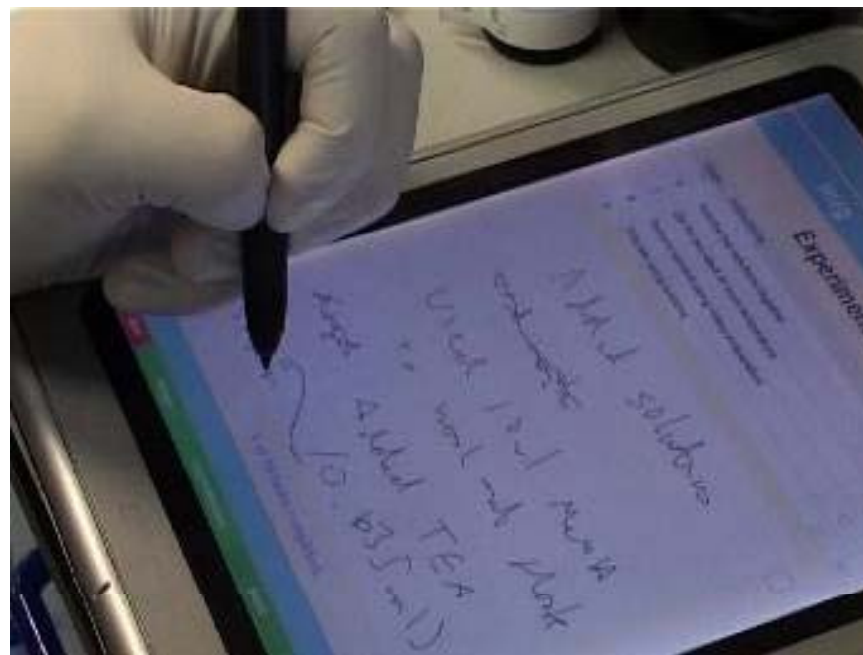


Beispiel: neue Formen der Datenerhebung 1

- Digitale Laborbücher in der Chemie



Quelle: Tony Hey, Microsoft Research



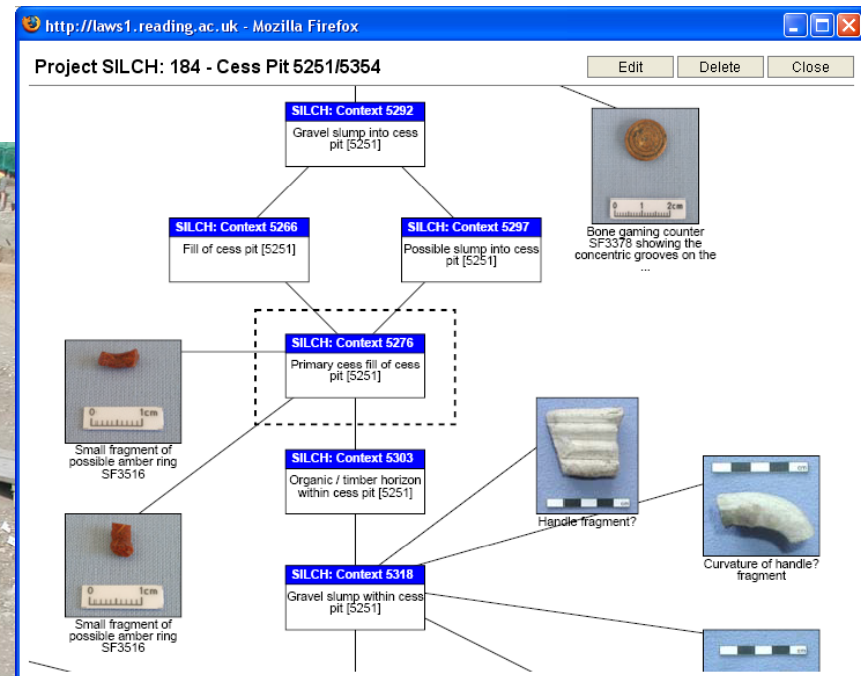
Quelle: Tony Hey, Microsoft Research

Beispiel: neue Formen der Datenerhebung 1

- Datenerhebung in der Archäologie



Quelle: Silchester Town Life Project; <http://www.silchester.rdg.ac.uk/home>



Conducting a Data Interview



... 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. 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.

<http://www.dcc.ac.uk/events/dcc-2007/programme/all-posters-demos>

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)
Assistant Professor of Library Science

Jake Carlson (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

Skills, role and career structure of data scientists and data curators

Eine Studie von KEY PERSPECTIVES für JISC, Juli 2008

- **Data creator**
Researchers with domain expertise who produce data. These people may have a high level of expertise in handling, manipulating and using data.
- **Data scientist**
People who work where the research is carried out – or, in the case of data centre personnel, in close collaboration with the creators of the data – and may be involved in creative enquiry and analysis, enabling others to work with digital data, and developments in data base technology.
- **Data manager**
Computer scientists, information technologists or information scientists and who take responsibility for computing facilities, storage, continuing access and preservation of data.
- **Data librarian**
People originating from the library community, trained and specializing in the curation, preservation and archiving of data.

„Data Librarianship“ – Rollen, Aufgaben, Kompetenzen



Rat für Sozial- und
Wirtschaftsdaten (RatSWD)
www.ratswd.de

RatSWD

Working Paper Series

Working Paper

Nr. 144

„Data Librarianship“ –
Rollen, Aufgaben, Kompetenzen

Heinz Pampel, Roland Bertelmann
und Hans-Christoph Hobohm

Mai 2010



... Die vielfältigen und häufig disziplinspezifischen Herausforderungen beim Umgang mit wissenschaftlichen Daten fordern eine engere Kooperation zwischen Wissenschaft und infrastrukturellen Serviceeinrichtungen. Bibliotheken bietet sich die Chance, die Entwicklung organisatorischer und technischer Lösungen des Forschungsdatenmanagements aktiv zu gestalten und eine tragende Rolle in diesem Feld zu übernehmen. ...

98. Deutschen Bibliothekartag
2. bis 5. Juni 2009 in Erfurt

DFG

Das Programm, Teil 3

- Maßnahmen der DFG und anderer Wissenschaftsorganisationen



Quelle: DFG

Die Deutsche Forschungsgemeinschaft

- **1998: DFG Denkschrift: “Sicherung guter Wissenschaftlicher Praxis”**
Primärdaten als Grundlagen für Veröffentlichungen sollen auf haltbaren und gesicherten Trägern in der Institution, wo sie entstanden sind, für zehn Jahre aufbewahrt werden.
- **2003: Berliner Erklärung über den offenen Zugang zu wissenschaftlichem Wissen**
... Berliner Erklärung mit dem Ziel aufgesetzt, das Internet als Instrument für eine weltweite Basis wissenschaftlicher Kenntnisse und menschlicher Reflektion zu fördern und die erforderlichen Maßnahmen zu formulieren, die von Entscheidungsträgern, Forschungsorganisationen, Förderinstitutionen, Bibliotheken, Archiven und Museen zu bedenken ...
- **2006: DFG Positionspapier**
Wissenschaftliche Literaturversorgungs- und Informationssysteme -
Förderschwerpunkte bis 2015

Die Deutsche Forschungsgemeinschaft

- **seit 2007: regelmäßige Workshops und Expertenrundgespräche**
Wissenschaftler treffen mit Informationsmanagement Experten zusammen, Ausrichtung ist in der Regel Disziplinspezifisch
- **2008: Nationale Schwerpunktinitiative „Digitale Information“ der Allianz der Deutschen Wissenschaftsorganisationen**
Arbeitsgruppe zu Forschungsdaten: abgestimmte Maßnahmen, Grundsatzpapier
- **2010: GWK Kommission Zukunft der Informationsinfrastruktur („KII“)**
Arbeitsgruppe zu Forschungsdaten: Empfehlungen an die GWK

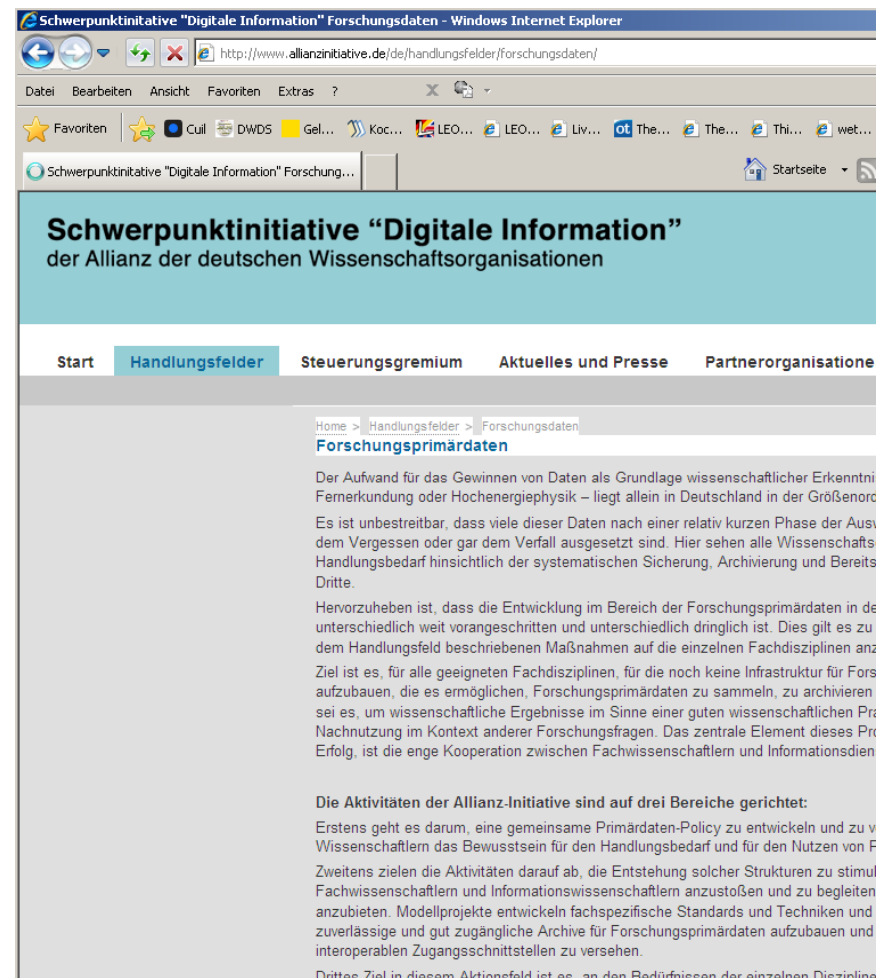
Die Allianz der Deutschen Wissenschaftsorganisationen

- **Die DFG gemeinsam mit den Allianzpartnern:**
Max-Planck-Gesellschaft
Helmholtz Gemeinschaft
Leibniz Gemeinschaft
Fraunhofer Gesellschaft
und weitere

- **22 Juni 2008: Schwerpunktinitiative "Digitale Information"**

- **Die Schwerpunkte sind:**
Nationale Lizenzierung
Open Access
Nationale Hosting Strategie
Forschungsdaten
Virtuelle Forschungsumgebungen
Rechtliche Rahmenbedingungen

Kolloquium Wissensinfrastruktur
Dr. Stefan Winkler-Nees; Bielefeld, 17. Dezember 2010



DFG Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsdaten

Ausschuss für Wissenschaftliche Bibliotheken und Informationssysteme
Unterausschuss für Informationsmanagement

Januar 2009

- Definitionen zu Forschungsdaten,
Organisationskonzept, Metadaten und Standards
- Rechtewahrung der Wissenschaftlerinnen und
Wissenschaftler
- Überregionale Bereitstellung
- Qualitätssicherung

www.dfg.de/lis/ unter „Veröffentlichungen / Informationsmanagement“

Aktuelle Aktivitäten der DFG

Ausschreibung: Informationsinfrastrukturen für Forschungsdaten

Die Deutsche Forschungsgemeinschaft (DFG) unterstützt mit dieser Ausschreibung im Förderbereich Wissenschaftliche Literaturversorgungs- und Informationssysteme (LIS) Vorhaben zur Entwicklung und Optimierung von Informationsinfrastruktur, die auf einen effizienten und nachhaltigen Umgang mit Forschungsdaten abzielen.

Ergänzung des allgemeinen Antragsmusters (auch in der neuen Runde der Exzellenzinitiative)

„ ... 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.“

Forschungsdaten: Ansätze und Strategien

- Begleitung eines Gestaltungsprozesses mit dem Ziel, digitale Ressourcen besser zu nutzen, Infrastrukturen aufzubauen und Werkzeuge zu entwickeln und Daten bereit zu stellen:

(Sicherung und Archivierung und Nachnutzung)

- Enge Einbindung der Wissenschaftler als Datenproduzenten und Nutzer der Datenrepositorien gemeinsam mit Experten aus dem Informationsmanagement in die Gestaltung dieses Prozesses.
- Fachspezifische Anforderungen und Bedürfnisse formulieren und in den Prozess mit einbringen.
- Entwicklung von Publikationsformen für wissenschaftlichen Daten (inkl. Peer-Review).
- Auf existierenden Ansätzen aufbauen und von Erfahrungen profitieren.
- Pilotprojekte und Explorationsprojekte initiieren.

*Die DFG fördert die systematische
Zusammenarbeit zwischen
Wissenschaftlern und
Informationsspezialisten.*

Zukünftige Aktivitäten

- Vernetzung nationaler und internationaler Informationsinfrastrukturen – Interoperabilität
- Neue Fördermöglichkeiten für Informationsinfrastrukturen
- Juristische Themen
- Disziplinspezifische Workshops und Rundgespräche
- Digitale Sammlungen wissenschaftlicher Objekte

Digitale Sammlungen wissenschaftlicher Objekte: Forschungsdaten?

PI für museale Objekte: inetbib, 15. Dezember 2010

[...] Mitglieder der ICOM/CIDOC Arbeitsgruppe "Transdisciplinary Approaches in Documentation" haben nun einen Vorschlag zu eindeutigen und persistenten Identifikatoren für Museumsobjekte erarbeitet. Der Vorschlag definiert die Rahmenbedingungen und die Syntax sog. "Museum Object Identifier" (MOI), die für den Einsatz beim Datenaustausch gedacht sind. Er berücksichtigt dabei die Rahmenbedingungen der Europeana und ähnlicher Portale sowie die Anforderungen, die sich durch den Einsatz von Technologien aus dem Umfeld des Semantic Web ergeben.

Eine erste Version von MuseumID steht ab sofort online zur Verfügung:

<http://museumid.net/documentation>

Anmerkungen, Fragen und Kritik über die Liste oder via contact@museumid.net sind herzlich willkommen.

Mit freundlichen Grüßen,
Georg Hohmann

--

Germanisches Nationalmuseum
Referat für Museums- und Kulturinformatik

Das Programm, Teil 4

- Ein Fazit



Quelle: DFG

Was ist zu tun? Eine Zusammenfassung

▶ Zugang

- Formen und Bedingungen des Zugangs regeln (“Open Access?”)

▶ Unterschiede in den Disziplinen

- Art der Daten, Menge und Typus, Lebenszyklen und Nutzungscharakteristik definieren

▶ Wissenschaftliche Anerkennung – Anreizsysteme

- Die Bereitstellung von Daten für die Nachnutzung als selbstverständlichen Bestandteil wissenschaftlichen Arbeitens und Anerkennung etablieren (Publikationskultur)

▶ Gebrauch von Standards, Entwicklung und Implementierung von Infrastruktur

- Zusammenarbeit zwischen Wissenschaftlern und Informationsmanagementexperten
- Sicherung der Interoperabilität in internationalen und interdisziplinären Netzwerken

Forschungsdatenmanagement an der Universität?

Stichpunkte für die Diskussion ...

▶ **Kein Projekt ohne Datenmanagementplan**

- Wie, was, wer

▶ **Anleitung zum Umgang mit Forschungsdaten**

- Der professionelle Umgang mit Forschungsdaten muss selbstverständlicher Bestandteil wissenschaftlichen Arbeitens werden.

▶ **Lehre und Qualifikation**

- Wissenschaftliche Institutionen, Forschende und Informationsdienstleister
- Einbindung der wissenschaftlichen Bibliotheken und Archive, sowie der Rechenzentren

A blurred, 3D-style logo of the German Research Foundation (DFG) in white, set against a blue background with light rays.

Vielen Dank für die Aufmerksamkeit!

Stefan.Winkler-Nees@dfg.de