



Digitale Langzeitarchivierung an Hochschulen

Forschungsdaten - Quo vadis

Prof. Dr. Peter Schirmbacher

Humboldt-Universität zu Berlin

Institut für Bibliotheks- und Informationswissenschaft

Computer- und Medienservice

Berlin, 29.04.2013



Forschungsdaten - Quo vadis



1. Einführung



Quelle: <http://aviation-research.de/assets/images/forschung.jpg>



29.04.2013

1. Was sind Forschungsdaten?

Forschungsprimärdaten

sind Daten, die im Verlauf von Quellenforschungen, Experimenten, Messungen, Erhebungen oder Umfragen entstanden sind. Sie stellen die Grundlagen für die wissenschaftlichen Publikationen dar. In Abhängigkeit von der jeweiligen Fachzugehörigkeit sind die Forschungsprimärdaten unterschiedlich zu definieren. Die Wissenschaftler sollen in ihren Fachcommunities selber entscheiden, ob bereits Rohdaten hierzu zählen oder **ab welchem Grad der Aggregation** die Daten langfristig aufzubewahren sind.

[Quelle: DFG, 2009]

1. Was sind Forschungsdaten?

„Welche Daten sollen und müssen archiviert werden? Die Empfehlungen der DFG sprechen von **Forschungsprimärdaten**. Dieser Begriff „Primärdaten“ sorgt immer wieder für Diskussion, denn die Definition des Begriffs ist sehr von der eigenen Rolle in der **wissenschaftlichen Wertschöpfungskette** bestimmt.

Für den einen sind „Primärdaten“ der Datenstrom aus einem Gerät, z.B. einem Satelliten. In der Fernerkundung werden diese Daten „Level 0“ Produkte genannt. Für andere sind „Primärdaten“ zur Nachnutzung aufbereitete Daten, ohne weiterführende Prozessierungsschritte. Wieder andere differenzieren nicht nach Grad der Verarbeitung sondern betrachten alle Daten, die Grundlage einer wissenschaftlichen Veröffentlichung waren, als Primärdaten. **Der begrifflichen Klarheit wegen sollte daher das Präfix „Primär-“ nicht mehr verwendet werden und statt dessen nur noch von wissenschaftlichen Daten oder Forschungsdaten gesprochen werden.**“

[Quelle: Klump, 2009, 105]

Forschungsdaten - Quo vadis



1. Was sind Forschungsdaten?

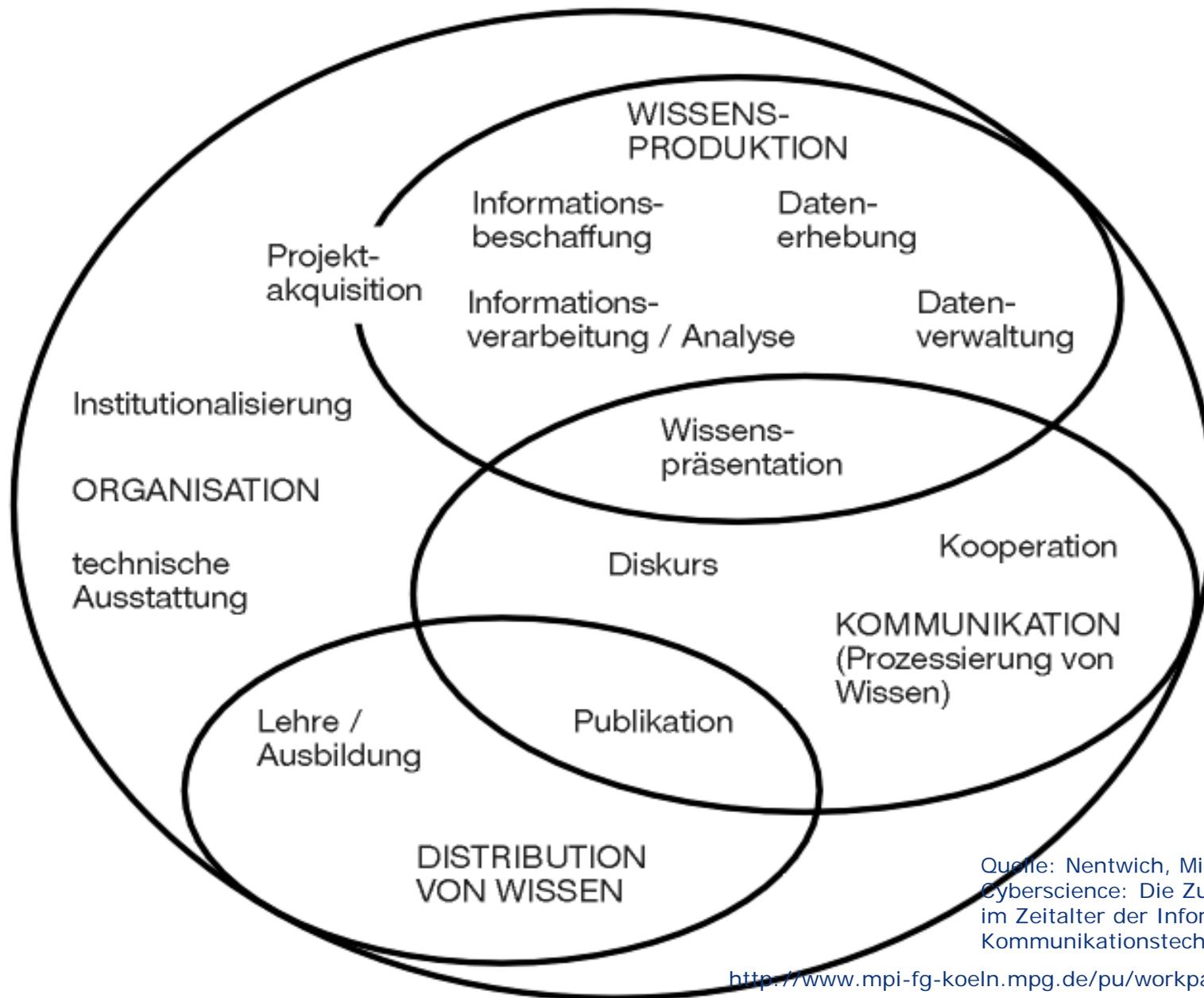
„Unter digitalen Forschungsdaten verstehen wir [...] alle digital vorliegenden Daten, die während des Forschungsprozesses entstehen oder ihr Ergebnis sind. Der Forschungsprozess umfasst dabei den gesamten Kreislauf von der Forschungsdatengenerierung, z. B. durch ein Experiment in den Naturwissenschaften, eine dokumentierte Beobachtung in einer Kulturwissenschaft oder eine empirische Studie in den Sozialwissenschaften, über die Bearbeitung und Analyse bis hin zur Publikation und Archivierung von Forschungsdaten.“

Digitale Forschungsdaten entstehen in allen Wissenschaftsdisziplinen und unter Anwendung verschiedener Methoden, abhängig von der Forschungsfrage. Dies hat zur Folge, dass sie in unterschiedlichen Medientypen, Aggregationsstufen und Datenformaten auftreten. Um das Bereitstellen von Forschungsdaten und ihre Nachnutzung zu ermöglichen, sind Metadaten und eine Datendokumentation essentiell, die den Kontext der Forschungsdaten beschreiben sowie die Werkzeuge, mit denen sie erzeugt, gespeichert, bearbeitet und analysiert wurden.“

[Quelle: Kindling, Maxi & Schirnbacher, Peter (2013) Die digitale Forschungswelt als Gegenstand der Forschung. In: Information, Wissenschaft & Praxis 2-3/2013, erscheint]



Schaubild 1 Die wissenschaftlichen Aktivitätsformen und Rahmenbedingungen



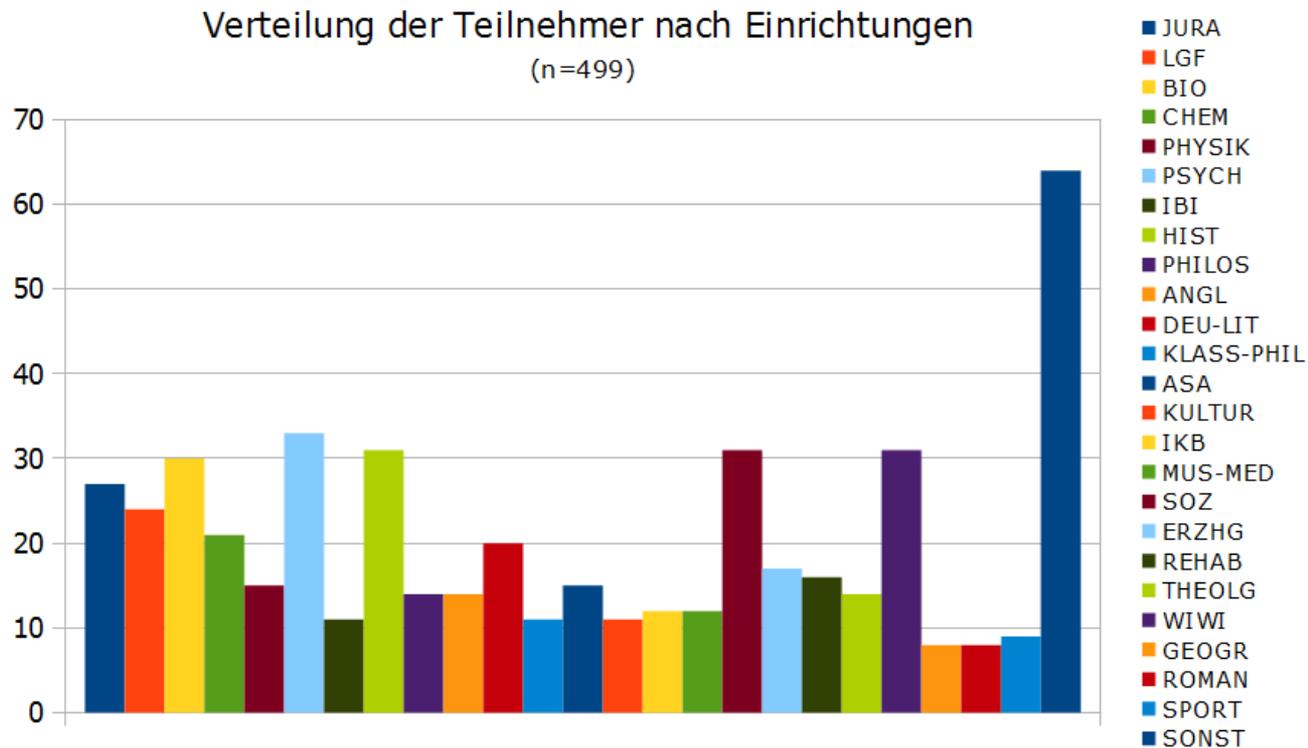
Quelle: Nentwich, Michael;
Cyberscience: Die Zukunft der Wissenschaft
im Zeitalter der Informations- und
Kommunikationstechnologien

1. Was sind Forschungsdaten?

- Der Begriffsumfang ist im Grunde unbestimmt.
- Das Verständnis von digitalen Forschungsdaten ist abhängig von verschiedenen Aspekten wie:
 - Wissenschaftliche Disziplinen
 - Wissenschafts„kultur“
 - Methoden & Werkzeuge in Abhängigkeit von Forschungsfragen
 - Formen und Formate der Daten
 - Aggregationsstufen von Daten

1. Was sind Forschungsdaten?

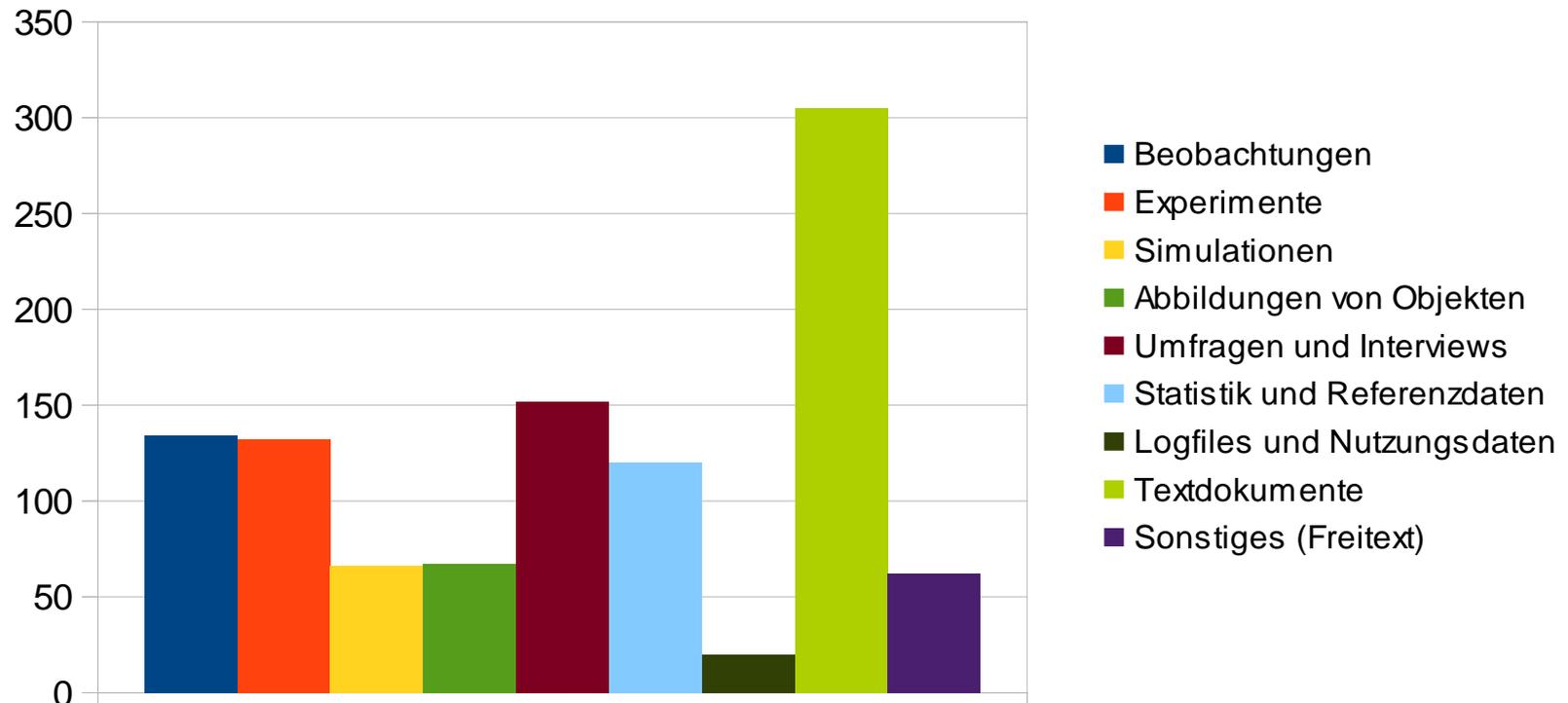
Umfrage an der HU zum Umgang mit Forschungsdaten / 1. Quartal 2013



Quelle: Umfrage in der HU im 1. Quartal 2013
Elena Simukovic, Maxi Kindling, Peter Schirmbacher

1. Was sind Forschungsdaten?

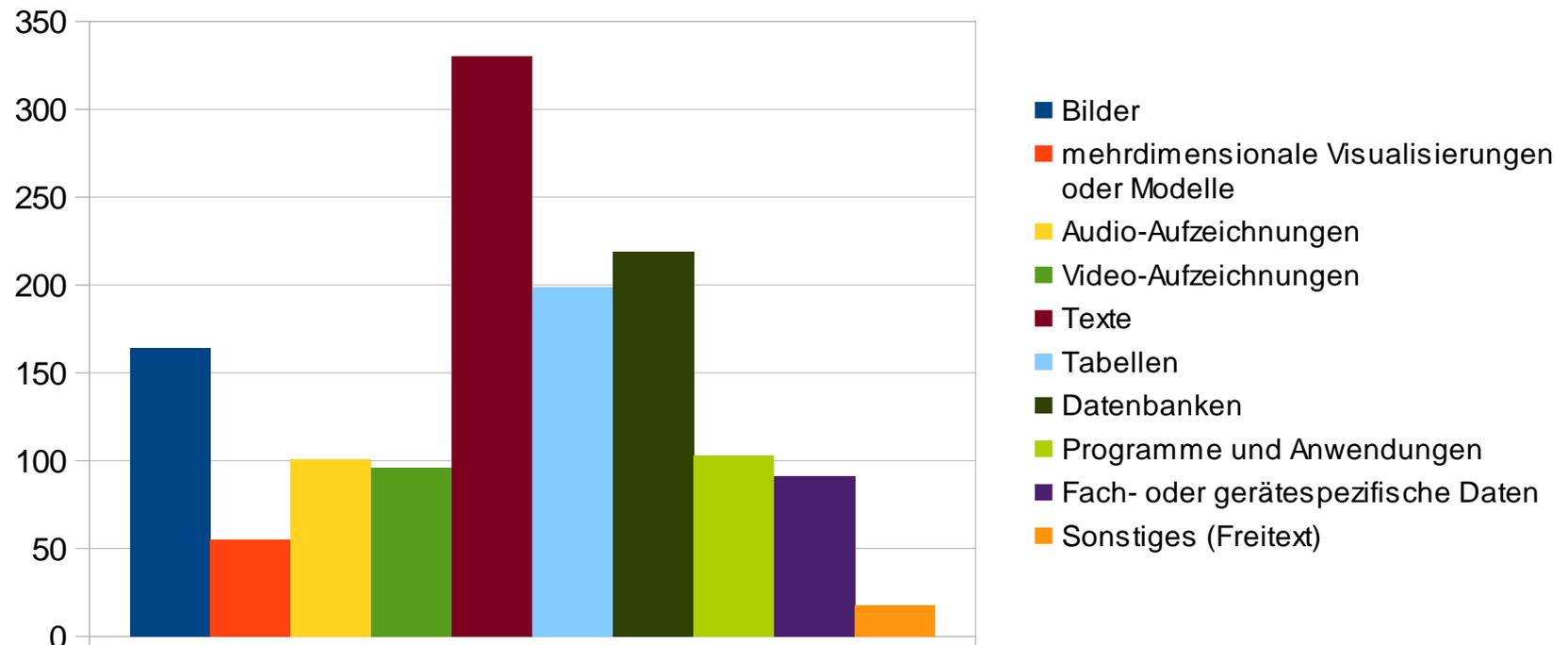
Aus welchen Quellen schöpfen Sie Ihre Forschungsdaten hauptsächlich?



Quelle: Umfrage in der HU im 1. Quartal 2013
Elena Simukovic, Maxi Kindling, Peter Schirmbacher

1. Was sind Forschungsdaten?

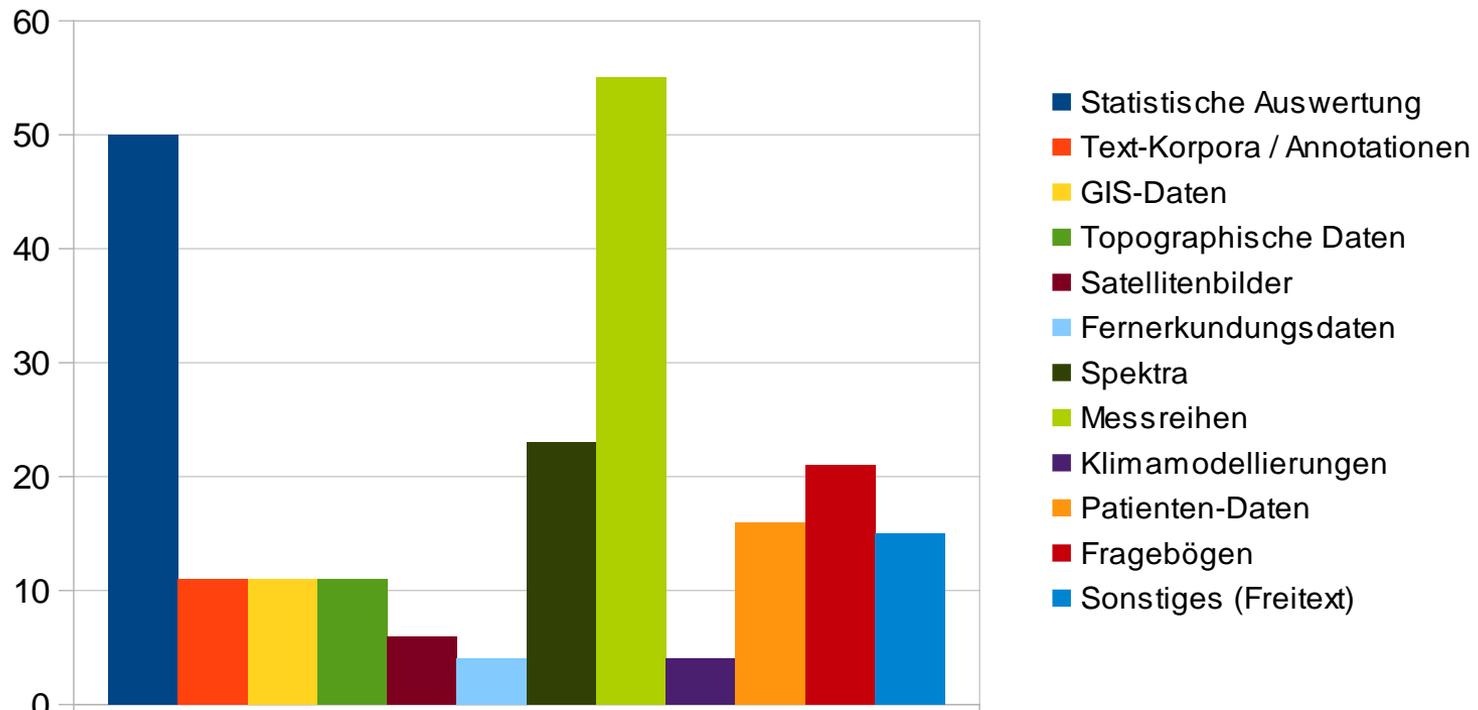
Um welche Datentypen handelt es sich?



Quelle: Umfrage in der HU im 1. Quartal 2013
Elena Simukovic, Maxi Kindling, Peter Schirmbacher

1. Was sind Forschungsdaten?

Um welche fach- oder gerätespezifischen Daten handelt es sich?



Quelle: Umfrage in der HU im 1. Quartal 2013
Elena Simukovic, Maxi Kindling, Peter Schirmbacher

Funktionen von Forschungsdaten

- Nachweisinstrument
- Kommunikation
- Reputation
- Nachnutzung

3. u. 4. Wann sollten Forschungsdaten aufbewahrt werden und wer sollte dies tun?

Der Umgang mit Forschungsdaten hat mindestens drei Dimensionen:

Wissenschaftspolitische Dimension

Wer ist verantwortlich und wie sehen die politischen, finanziellen und personellen Rahmenbedingungen regional, national und international aus.

Organisatorische Dimension

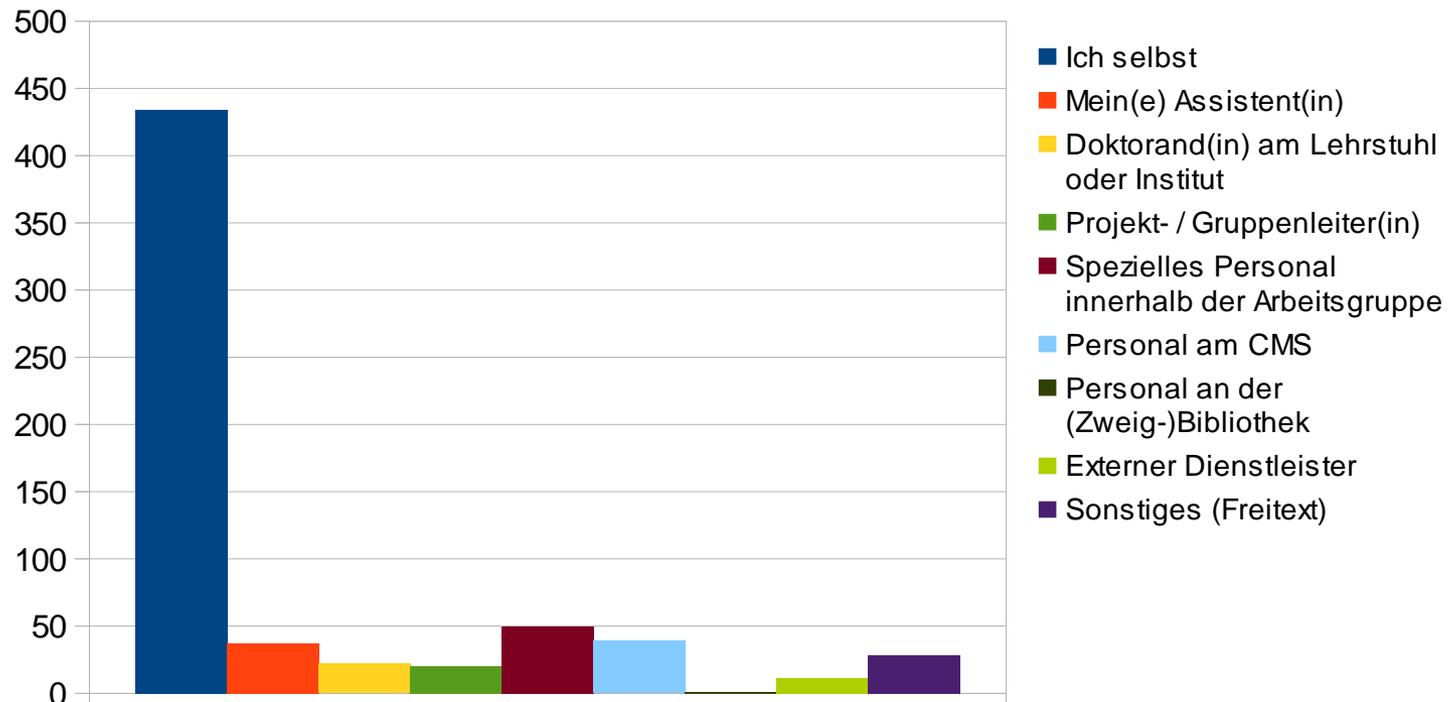
Welche Strukturen werden geschaffen und wie kann eine entsprechende Arbeitsteilung zwischen den Einrichtungen und zwischen Wissenschaftlern und Servicepersonal aussehen.

Technische Dimension

Auf welcher technischen Basis wird die Infrastruktur geschaffen. Wie wird ein entsprechender Service unter Einhaltung sämtlicher rechtlichen Rahmenbedingungen gewährleistet.

3. u. 4. Wann sollten Forschungsdaten aufbewahrt werden und wer sollte dies tun?

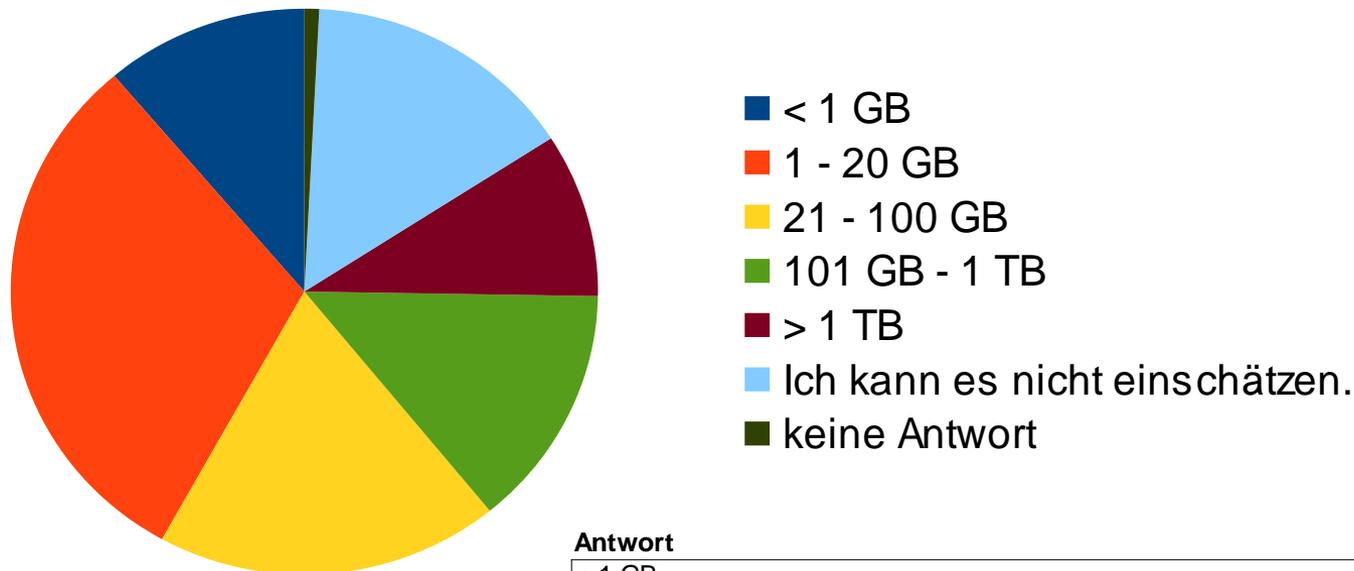
Wer ist verantwortlich für die Speicherung, Sicherung oder Archivierung Ihrer Forschungsdaten?



Quelle: Umfrage in der HU im 1. Quartal 2013
Elena Simukovic, Maxi Kindling, Peter Schirmbacher

3. u. 4. Wann sollten Forschungsdaten aufbewahrt werden und wer sollte dies tun?

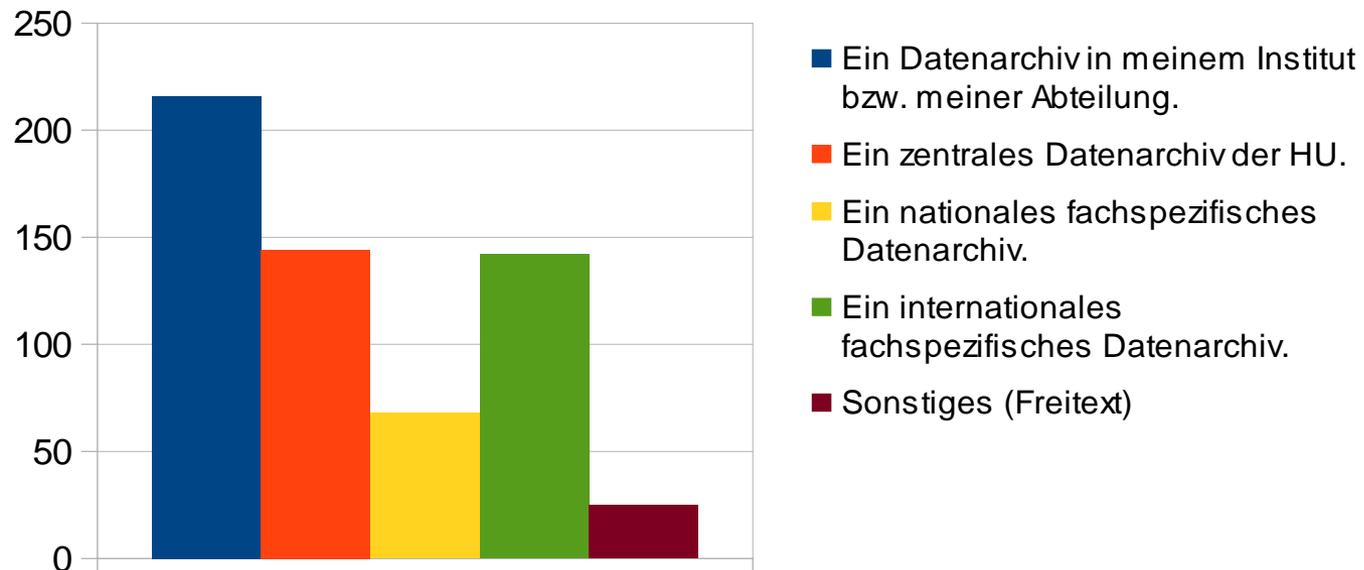
Bitte schätzen Sie die Gesamtgröße Ihrer Forschungsdaten bezogen auf den benötigten Speicherplatz.



Antwort	Anzahl	Prozent
< 1 GB	55	11.20%
1 - 20 GB	151	30.75%
21 - 100 GB	93	18.94%
101 GB - 1 TB	68	13.85%
> 1 TB	46	9.37%
Ich kann es nicht einschätzen.	74	15.07%
keine Antwort	4	0.81%

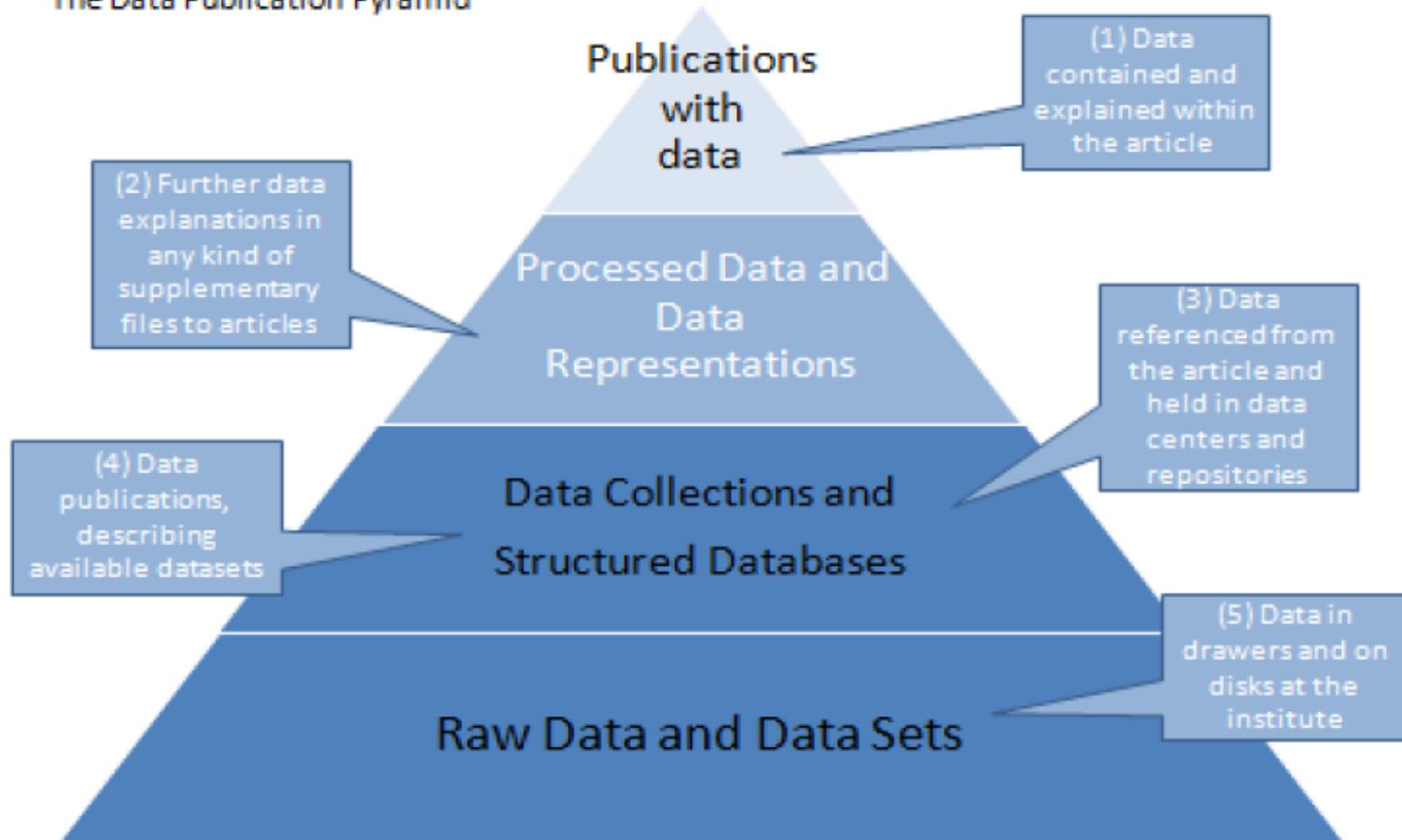
5. Wo sollten Forschungsdaten aufbewahrt werden?

Welches Datenarchiv würden Sie für die Ablage ihrer Daten am ehesten nutzen?



Quelle: Umfrage in der HU im 1. Quartal 2013
Elena Simukovic, Maxi Kindling, Peter Schirmbacher

The Data Publication Pyramid



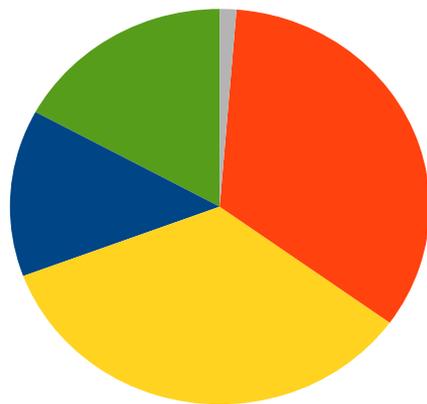
Graph 1: The Data Publication Pyramid, developed on the basis of the Jim Gray pyramid, to express the different manifestation forms that research data can have in the publication process.

See Chapter 1 for a full explanation.

Quelle: <http://oa.helmholtz.de/index.php?id=258>

5. Wo sollten Forschungsdaten aufbewahrt werden?

Haben Sie schon einmal Ihre Forschungsdaten in einem Datenarchiv abgelegt?



- Ja.
- Nein, aber ich habe es vor.
- Nein, diese Möglichkeit war mir nicht bekannt.
- Nein, in der nächsten Zeit habe ich es nicht vor.
- keine Antwort

Ja.	127	27.08%
Nein, aber ich habe es vor.	48	10.23%
Nein, diese Möglichkeit war mir nicht bekannt.	154	32.84%
Nein, in der nächsten Zeit habe ich es nicht vor.	131	27.93%
keine Antwort	9	1.92%

re3data.org

REGISTRY OF RESEARCH DATA REPOSITORIES

[Home](#) [Search](#) [Suggest](#) [FAQ](#) [About](#) [Schema](#) [Contact](#) [Imprint](#)

Search for repositories (alpha version)

Filter results

Search Results (173 results)

CLARIN



Linguistics

CLARIN aims at uniting existing digital archives in Europe that contain language-based material into a federation that will allow the social sciences and humanities research communities to have unified access to the content. It wants to make the wealth of language and speech processing tools that have been developed over the recent years available to interested researchers with a view to opening up new research avenues. Another goal of CLARIN is to provide web based services that will allow non-expert users (especially humanities and social sciences researchers without technological background) to perform complex tasks on the materials contained in the archives, such as 'Summarize Le Monde of March 17 2008 - in Polish'. Only countries and institutions can be CLARIN members. Private persons are not allowed.

DANS



Ancient Cultures Geosciences (including Geography) History Social and Behavioural Sciences

DANS promotes sustained access to digital research data. For this purpose, DANS encourages researchers to archive and reuse data in a sustained manner, e.g. through the online archiving system EASY. DANS also provides access, via NARCIS.nl, to thousands of scientific datasets, e-publications and other research information in the Netherlands. In addition, the Institute provides training and advice, and performs research into sustained access to digital information. Driven by data, DANS ensures that access to digital research data keeps improving, through its services and by taking part in national & international projects and networks. DANS is an Institute of the Royal Netherlands Academy of Arts and Sciences (KNAW) and the Netherlands Organisation for Scientific Research (NWO).

Ecological Society of America esa Ecological Archives



Agriculture Biology Chemistry Forestry Horticulture and Veterinary Medicine Life Sciences

Reviewed Results

Search for repositories (alpha version)

mineralogy Search

Filter results »

Subject: Content Type: Country:

Certificates Open Access Persistent Identifier Repository reviewed by re3data.org

remove filters

Search Results (4 results)

Search terms: *mineralogy*

Crystallography Open Database

Physics

Open-access collection of crystal structures of organic, inorganic, metal-organic compounds and minerals, excluding biopolymers

Inorganic Crystal Structure Database

Chemistry Geochemistry Materials Science and Engineering Mineralogy and Crystallography Physics

The most comprehensive database on fully determined inorganic crystal structures • Full structural data: cell parameters, atom positions for all entries, displacement parameters • Full bibliographic data: publication title, journal reference(s), author names • Full structure description: Structural formula, compositions, ANX formulae, structure types • High-quality data: extensive data evaluation and correction by senior experts • Web and PC based software solutions, data updated twice a year • 25+ years of serving the scientific community

PANGAEA

Biology Geochemistry Geodesy Geoinformatics Geology and Paleontology Geophysics Mineralogy and Crystallography Oceanography Remote Sensing

The Information system PANGAEA is operated as an Open Access library aimed at archiving, publishing and distributing georeferenced data from earth system research. The system guarantees long-term availability of its content through a commitment of the operating institutions.



6. Wie sollten Forschungsdaten aufbewahrt werden?

Anforderungen an die Aufbewahrung wissenschaftlicher Daten

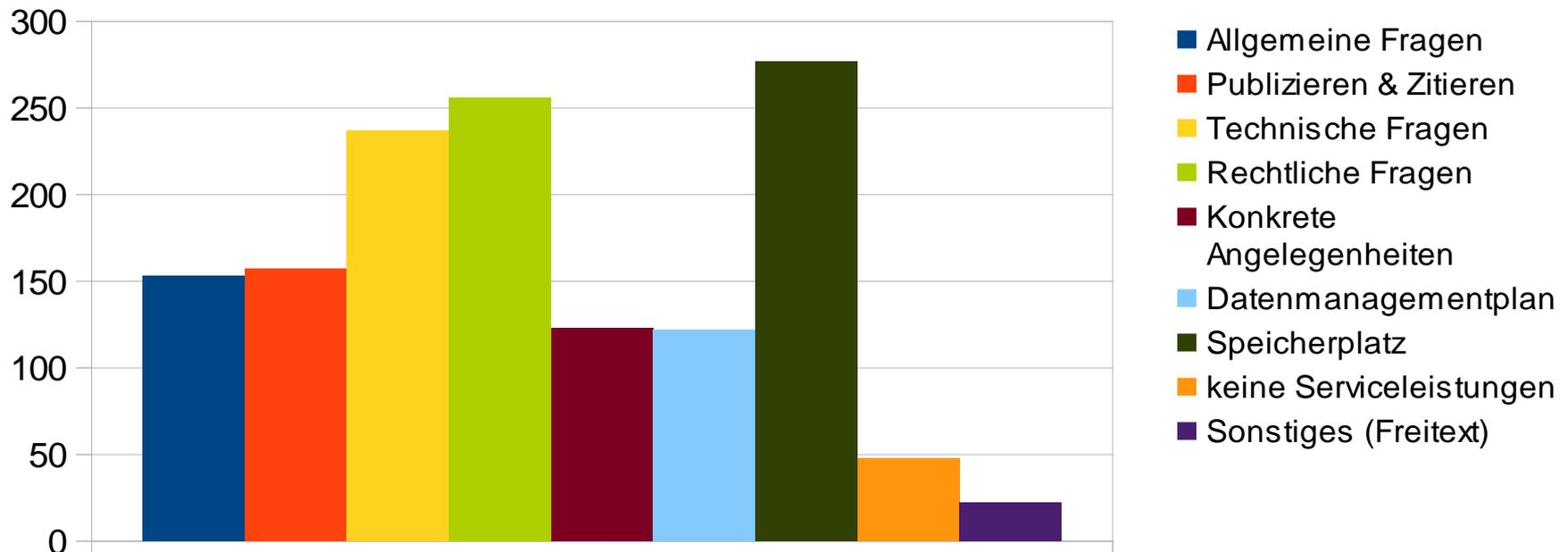
1. Zugänglichkeit
2. Nachhaltigkeit
3. Nachvollziehbarkeit
4. Authentizität
5. Qualitätssicherung
6. Bewertung
7. Geschwindigkeit
8. Vollständigkeit

Langzeitarchivierung von Forschungsdaten

- Was verstehen wir darunter?
- Welche Notwendigkeit gibt es?
- Welche Anforderungen werden gesehen?
- Welche Organisationsformen gibt es?
- Welche technischen Voraussetzungen sind zu schaffen?

5. Wie sollten Forschungsdaten aufbewahrt werden?

Welche Serviceleistungen würden Sie sich von der HU wünschen? (Forschungsdatenmanagement)



Quelle: Umfrage in der HU im 1. Quartal 2013
Elena Simukovic, Maxi Kindling, Peter Schirmbacher



Herzlichen Dank für die Aufmerksamkeit

???

Prof. Dr. Peter Schirmbacher

Humboldt-Universität zu Berlin

Institut für Bibliotheks- und Informationswissenschaft

Computer- und Medienservice

schirmbacher@cms.hu-berlin.de



29.04.2013