

“It's all about data.”

---



# Digitale Forschungsdaten an der Humboldt-Universität zu Berlin

Maxi Kindling, Peter Schirnbacher, Elena Simukovic  
Humboldt-Universität zu Berlin  
Institut für Bibliotheks- und Informationswissenschaft  
Computer- und Medienservice

Berliner Bibliothekswissenschaftliches Kolloquium, 21.05.2013



21.05.2013

# “It's all about data.”

---



Stand und Entwicklung im Umgang mit digitalen Forschungsdaten an der HU Berlin

## Agenda

- Einleitung –  
*Peter Schirmbacher*
- Projektmodul „Umgang mit digitalen FD an der HU Berlin“ –  
*Maxi Kindling*
- Umfrage zum Umgang mit digitalen Forschungsdaten an der HU Berlin –  
*Elena Simukovic*
- Diskussion

# “It's all about data.”

---



Was sind Forschungsdaten?

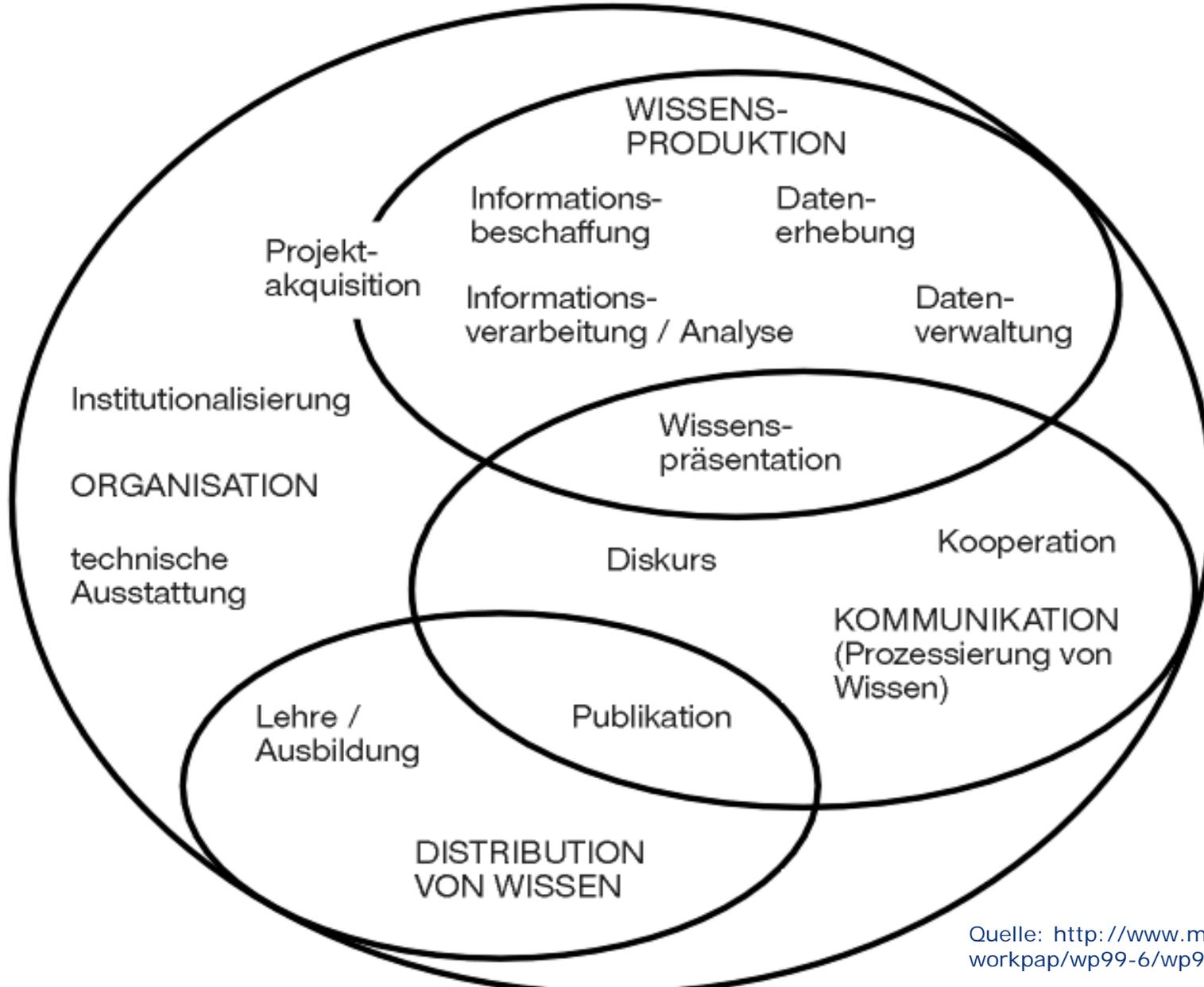
„Unter digitalen Forschungsdaten verstehen wir [...] alle digital vorliegenden Daten, die während des Forschungsprozesses entstehen oder ihr Ergebnis sind. Der Forschungsprozess umfasst dabei den gesamten Kreislauf von der Forschungsdatengenerierung, z. B. durch ein Experiment in den Naturwissenschaften, 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.“

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



21.05.2013

# Schaubild 1 Die wissenschaftlichen Aktivitätsformen und Rahmenbedingungen



# “It's all about data.”

---



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

# “It's all about data.”

---



Warum sollten Forschungsdaten aufbewahrt werden?

## Funktionen von Forschungsdaten

- Nachweisinstrument
- Kommunikation
- Reputation
- Nachnutzung

# “It's all about data.”

---



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

“It's all about data.”

---



# Projektmodul „Der Umgang mit digitalen Forschungsdaten an der HU Berlin“

Maxi Kindling

Humboldt-Universität zu Berlin

Institut für Bibliotheks- und  
Informationswissenschaft



21.05.2013

Maxi Kindling, Peter Schirnbacher, Elena Simukovic  
Institut für Bibliotheks- und Informationswissenschaft  
Computer- und Medienservice

8

# “It's all about data.”

---



Projektmodul „Umgang mit digitalen FD an der HU Berlin“

## Was?

- Projektmodul im Masterstudium

## Wann?

- Wintersemester 2012/13

## Wer war dabei?

- Seminarleitung: Maxi Kindling
- 11 Master-Studierende: Marleen Burger, Lisa Liebenau, Claudia Lienhard, Svantje Lilienthal, Paulina Plewka, Svenia Pohlkamp, Kristin Reinhardt, Melanie Rügenhagen, Kathleen Schulz, Kendra Sticht, Marcel Walther
- Unterstützung durch: Elena Simukovic, Peter Schirnbacher

# “It's all about data.”

---



Projektmodul „Umgang mit digitalen FD an der HU Berlin“

## Ziele des Projektmoduls

- Einblick in ein aktuelles Forschungsfeld der Bibliotheks- und Informationswissenschaft
- Aktive Beteiligung an einer Konzeptentwicklung für die HU Berlin
- Was ist der Status quo in Deutschland und international?
- Infrastrukturelle Aspekte:
  - Wissenschafts- und förderpolitischer Rahmen
  - Organisation
  - Rechtliche Belange (in Ansätzen)
  - Technische Konzepte (in Ansätzen)

# “It's all about data.”

---

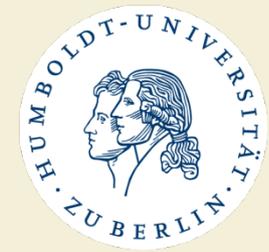


Projektmodul „Umgang mit digitalen FD an der HU Berlin“

## Was haben wir gemacht?

- Analyse exemplarischer deutscher und internationaler Ansätze des Forschungsdatenmanagements durch Gruppenreferate
  - Deutschland, Schweiz
  - UK, USA, Australien
- Umfangreiche Material- und Quellensammlung zum Thema
- Vorstellung der Ergebnisse in einem gemeinsamen Termin mit dem VPF
- Projektbericht mit Beteiligung aller Studierenden

# “It's all about data.”



Projektmodul „Umgang mit digitalen FD an der HU Berlin“



## Inhaltsverzeichnis

0 Vorwort.....	1
1 Einleitung.....	1
1.1 Inhalt, Zielgruppe und Motivation.....	1
1.2 Definition digitale Forschungsdaten.....	2
1.3 Aktualität des Themas.....	3
1.4 Forschungsdatenmanagement.....	3
2 Forschungsdatenmanagement an Hochschulen.....	4
2.1 Wissenschaftspolitische Vorgaben.....	4
2.2 Internationale Best Practices.....	6
Deutschland.....	6
Schweiz, UK, USA, Australien.....	10
3 Ausgewählte Aspekte eines zukünftigen Konzepts für das Forschungsmanagement an der Humboldt-Universität zu Berlin.....	13
3.1 Policies und Guidelines.....	13
3.2 Forschungsdatenmanagement-Pläne.....	15
3.3 Publikation der Forschungsdaten.....	16
3.4 Anforderungen an die Publikation und Archivierung digitaler Forschungsdaten.....	17
3.5 Szenarien für eine technische Infrastruktur zur Langzeitarchivierung und Publikation.....	18
3.5.1 Langzeitarchivierung und Publikation der Forschungsdaten.....	18
3.5.2 Speicherung während des Forschungsprozesses.....	22
3.5.3 Metadaten.....	23
3.6 Personelle Ausstattung und Aufgaben.....	23
3.7 Öffentlichkeitsarbeit.....	26
4 Zusammenfassung.....	28

## Forschungsdatenmanagement an Hochschulen

Internationaler Überblick und Aspekte eines  
Konzepts für die Humboldt-Universität zu Berlin



21.05.2013

# “It's all about data.”

---



Projektmodul „Umgang mit digitalen FD an der HU Berlin“

## Was sind Ergebnisse?

- disziplinübergreifender und serviceorientierter Blick auf das Forschungsdatenmanagement an Hochschulen
- Identifizierung von Best Practices
- UK, USA, Australien: große Bedeutung des FDM an Hochschulen
- deutsche Hochschulen im internationalen Vergleich noch am Anfang
- Auch: zukunftsweisende Projekte und Initiativen an Hochschulen aus Deutschland und der Schweiz (bspw. Uni Bielefeld, CAU Kiel, TU Berlin und ETH Zürich)

# “It's all about data.”

---



Projektmodul „Umgang mit digitalen FD an der HU Berlin“

## Was sind Empfehlungen?

- Policy zum adäquaten Umgang mit Forschungsdaten, deren Umsetzung konzeptionell internationalen Best Practices folgt
  - langfristige Archivierung (gemäß der Vorgaben der Wissenschaftspolitik und der Forschungsförderorganisation)
  - dauerhafte Zugänglichkeit für die Urheber, HU-Forschende, Öffentlichkeit (Open Access)
  - Nachweissystem
  - unterstützende Veranstaltungen
- technische Infrastrukturen, die sich in das Gesamtangebot der Service-Einrichtungen an der HU Berlin integrieren

# “It's all about data.”

---



Projektmodul „Umgang mit digitalen FD an der HU Berlin“

## Was sind Empfehlungen?

- Diskussion und Auswertung von Aspekten wie Policy, Forschungsdatenmanagement-Plan, Archivierung, Publikation, Aufgaben, personelle Ausstattung, Öffentlichkeitsarbeit gemeinsam mit den Wissenschaftsdisziplinen und interdisziplinären Projekten
- an der HU vorhandene Forschungs(daten)infrastrukturen und entsprechende Projekte integrieren
- Regionale, nationale und internationale Kooperationen

“It's all about data.”

---



# Umfrage zum Umgang mit digitalen Forschungsdaten an der HU Berlin

Elena Simukovic

Humboldt-Universität zu Berlin  
Computer- und Medienservice



21.05.2013

# “It's all about data.”

---



Umfrage zum Umgang mit digitalen Forschungsdaten an der HU

## Ziel & Aufbau

- Ziel: Ermittlung des Ist-Zustands und der Anforderungen an zukünftige Service-Angebote
- Zielgruppe: wissenschaftliches Personal der HU
- Fragebogen und Umsetzung baut auf vergleichbaren Umfragen an anderen Hochschulen (University of Cambridge, University of Glasgow, Imperial College London, ETH Zürich) und im PARSE.Insight-Projekt auf

# “It's all about data.”

---



Umfrage zum Umgang mit digitalen Forschungsdaten an der HU

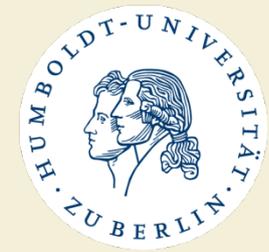
## Eckdaten

- 24 Fragen, davon 6 Pflichtfragen
- ca. 10 Minuten zum Ausfüllen
- Versionen in zwei Sprachen (Deutsch und Englisch)
- Online-Fragebogen (Nutzung des Open-Source-Produktes LimeSurvey)
- 6 Wochen Laufzeit (24.01.-08.03.2013)



21.05.2013

# “It's all about data.”



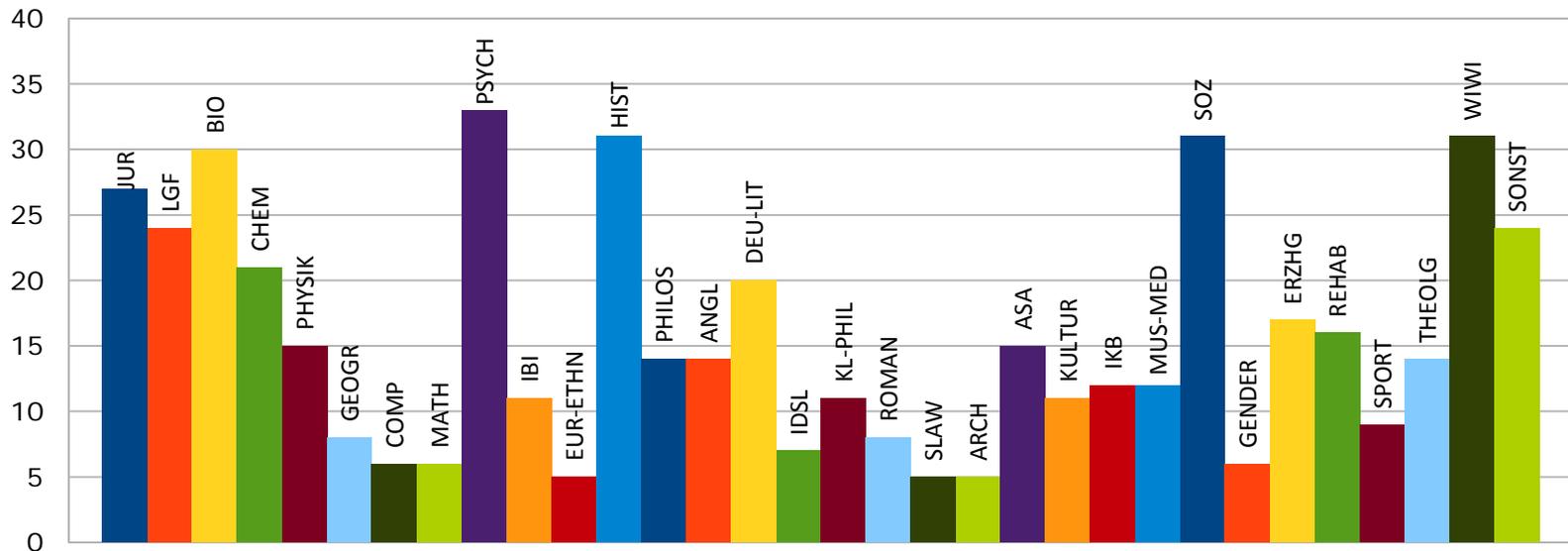
Umfrage zum Umgang mit digitalen Forschungsdaten an der HU

## Rücklauf

499 Antworten bzw. ~24% der Zielgruppe aus den meisten Einrichtungen:

Verteilung der Teilnehmer nach Einrichtungen

(n=499)



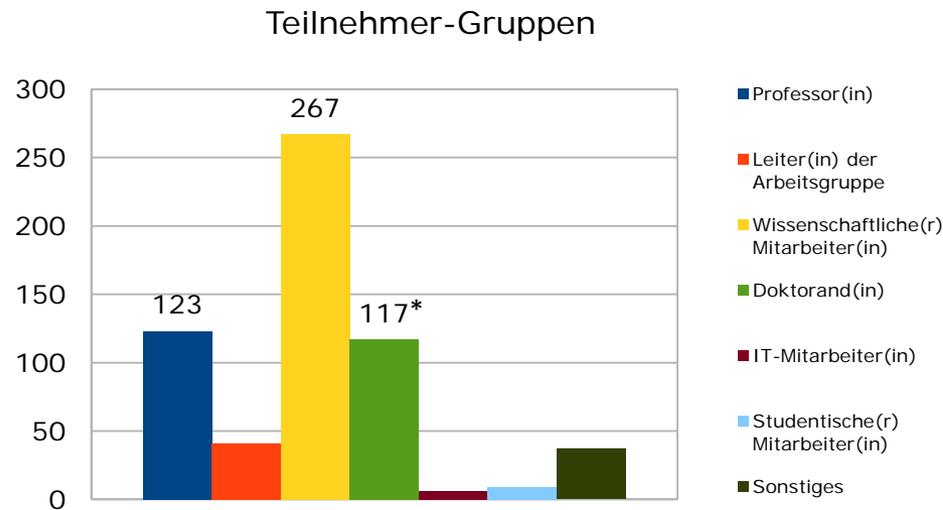
21.05.2013

# “It's all about data.”



Umfrage zum Umgang mit digitalen Forschungsdaten an der HU

## Teilnehmer-Gruppen



Entspricht etwa 29% der Professor/innen und 14% der wissenschaftlichen Mitarbeiter/innen der HU

Quelle: [HU-Personalstatistik](#) (Jahr 2013)

\*Anmerkung: 63 Teilnehmer haben als Doktorand(in) und wissenschaftliche(r) Mitarbeiter(in) geantwortet (Mehrfachantworten erlaubt)



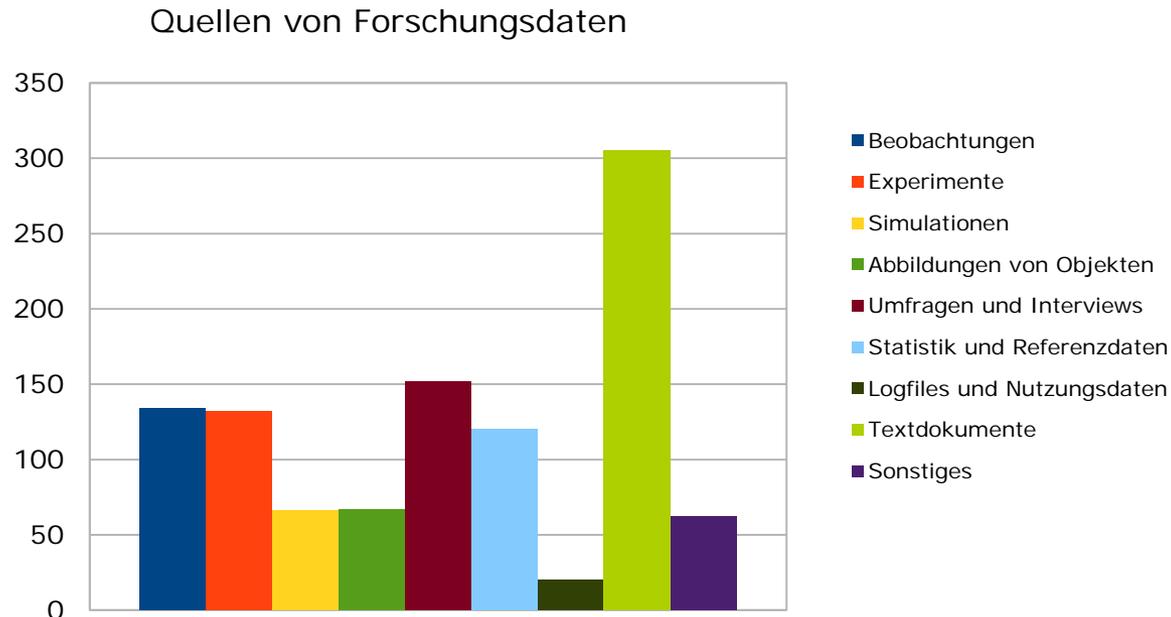
21.05.2013

# “It's all about data.”

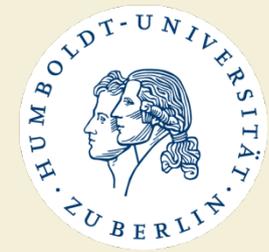


Umfrage zum Umgang mit digitalen Forschungsdaten an der HU

## Aus welchen Quellen schöpfen Sie Ihre Forschungsdaten hauptsächlich?

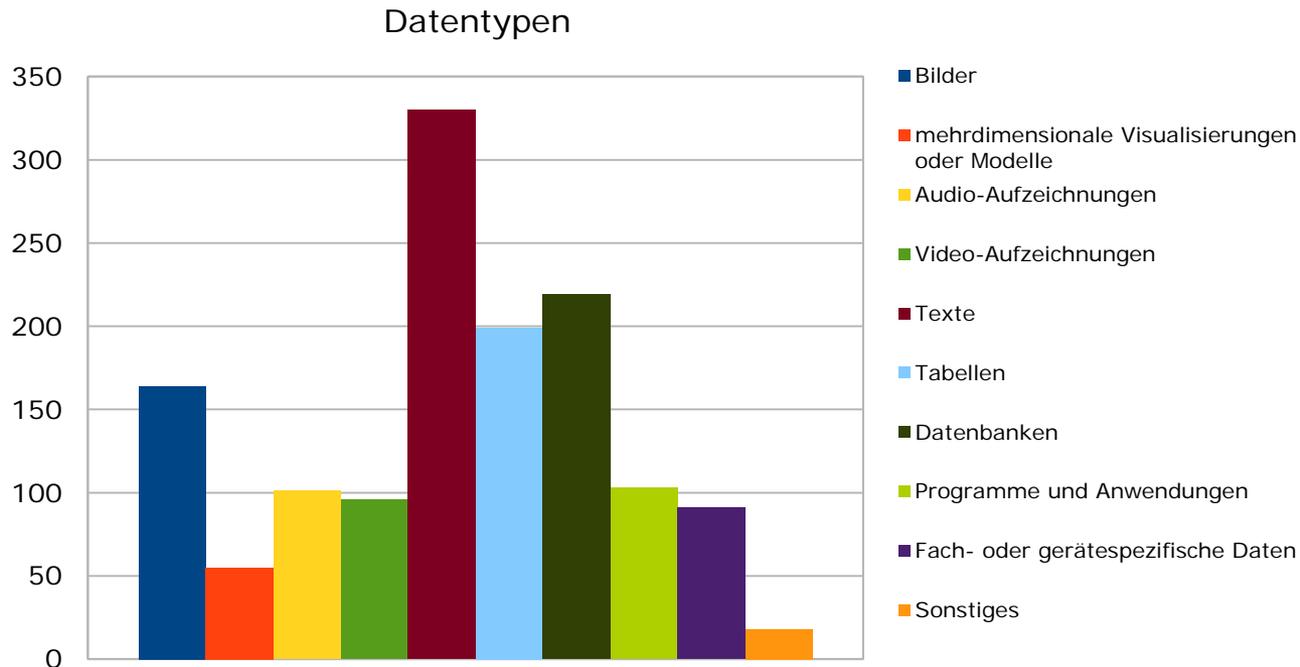


# “It's all about data.”



Umfrage zum Umgang mit digitalen Forschungsdaten an der HU

## Um welche Datentypen handelt es sich?



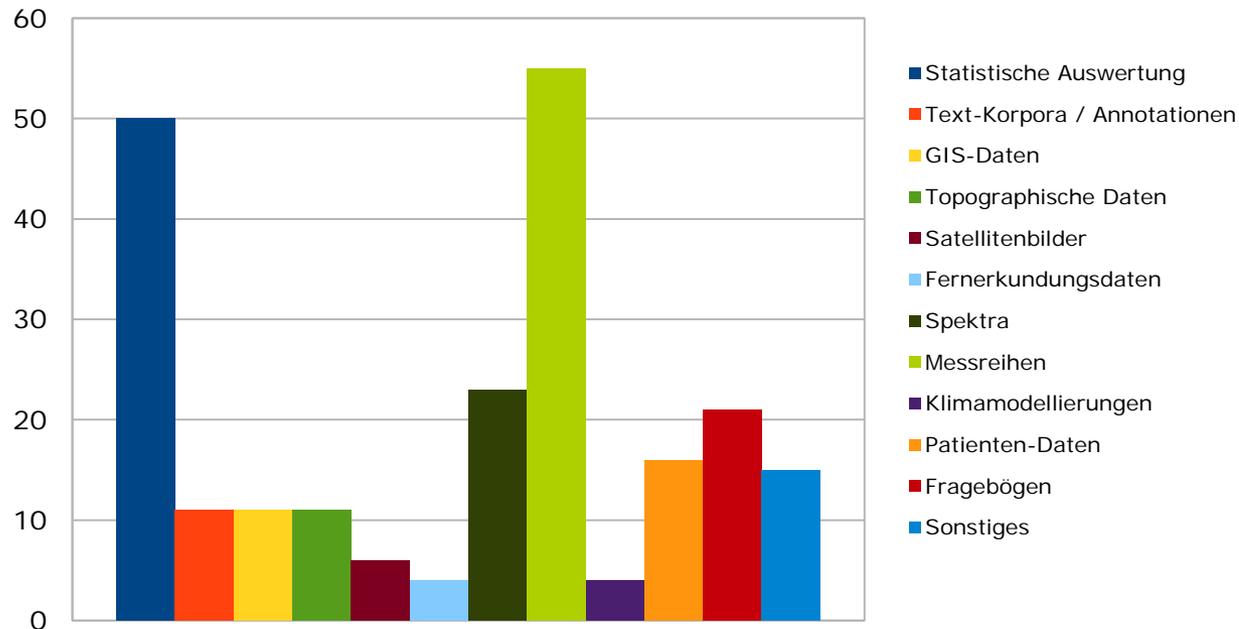
# “It's all about data.”



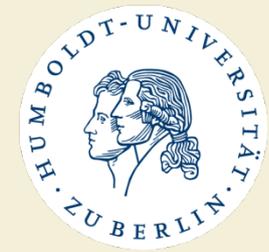
Umfrage zum Umgang mit digitalen Forschungsdaten an der HU

## Um welche fach- oder gerätespezifischen Daten handelt es sich?

Fach- oder gerätespezifische Daten

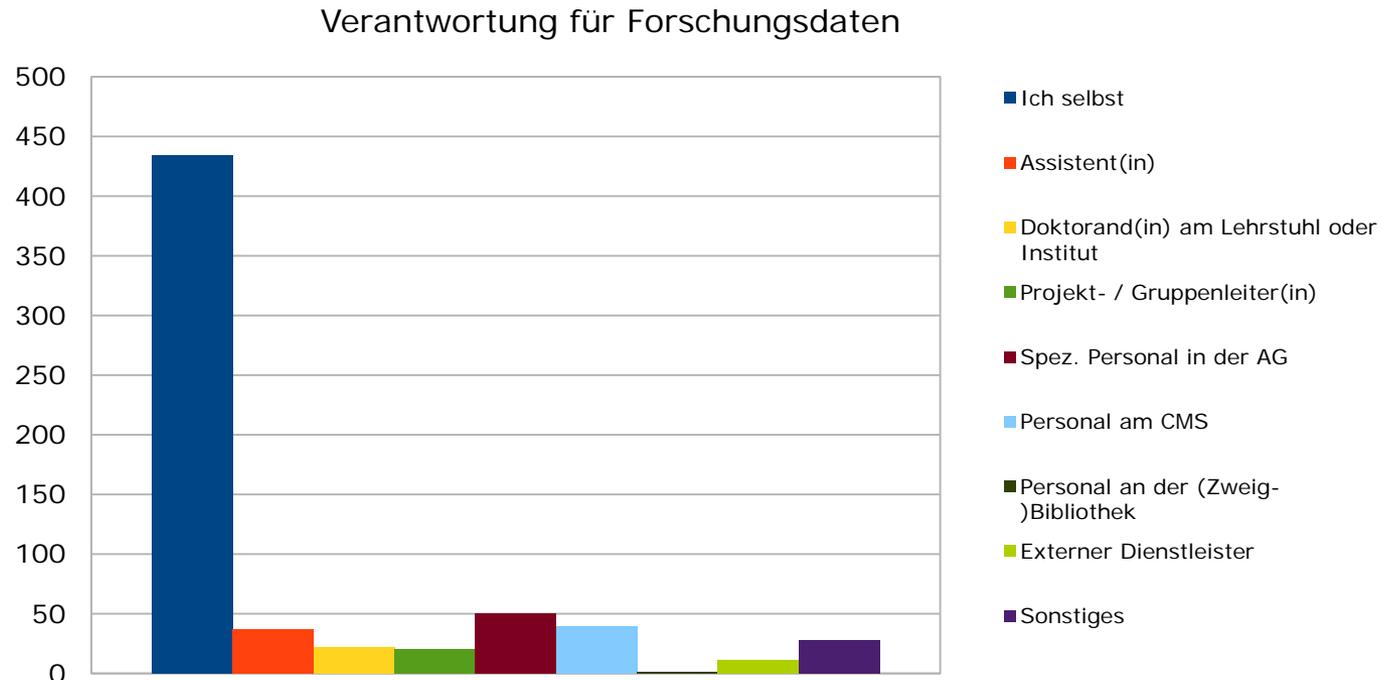


# “It's all about data.”



Umfrage zum Umgang mit digitalen Forschungsdaten an der HU

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



# “It's all about data.”

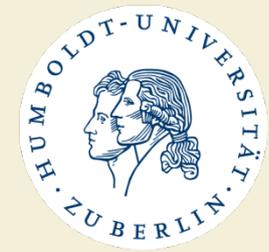


## Umfrage zum Umgang mit digitalen Forschungsdaten an der HU

In den Grundsätzen zur Sicherung guter wissenschaftlicher Praxis der HU (2002) heißt es: "Primärdaten als Grundlagen für Veröffentlichungen sollen auf haltbaren und gesicherten Trägern in der Arbeitsgruppe/ Einrichtung, wo sie entstanden sind, für zehn Jahre zugänglich bleiben." Berücksichtigen Sie diese Grundsätze?



# “It's all about data.”

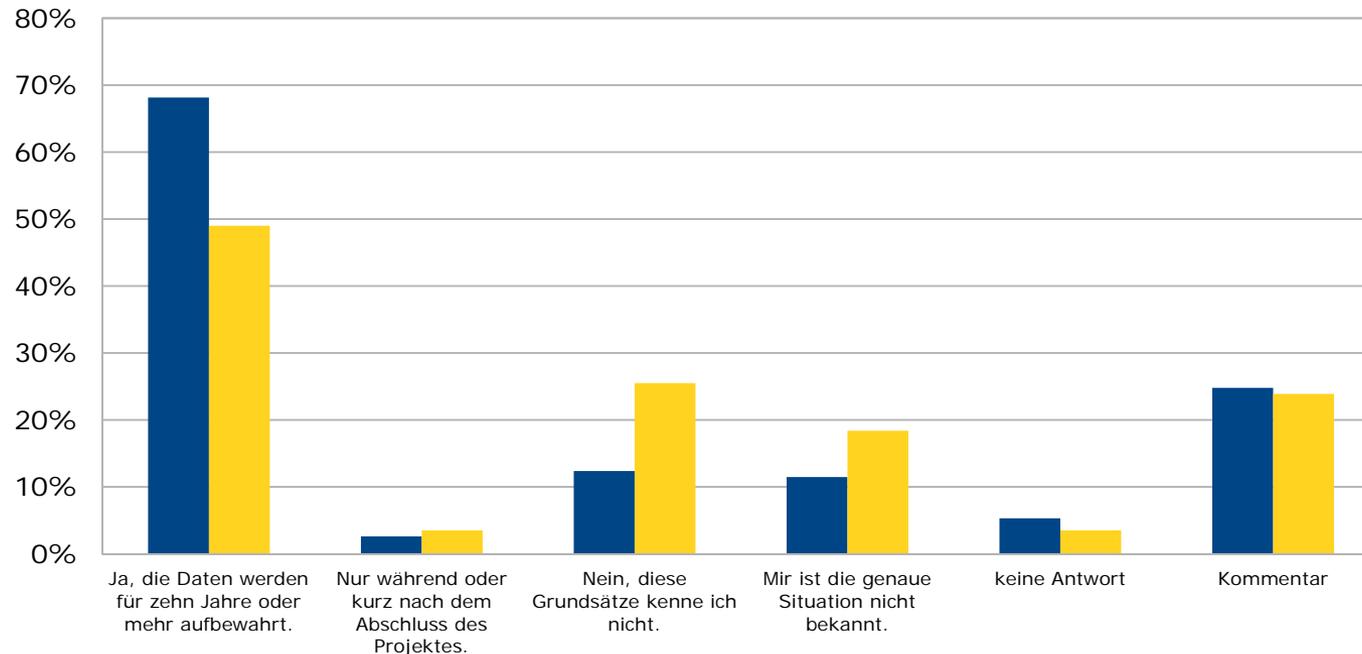


Umfrage zum Umgang mit digitalen Forschungsdaten an der HU

## Vergleich der Antworten zwischen Teilnehmer-Gruppen "Professor(in)" und "wissenschaftliche(r) Mitarbeiter(in)"

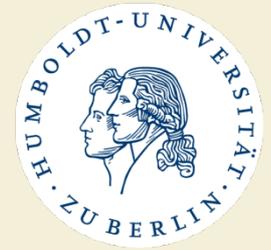
Sicherung guter wissenschaftlicher Praxis

■ Professor(in)    ■ wissenschaftliche(r) Mitarbeiter(in)



# “It's all about data.”

---



Umfrage zum Umgang mit digitalen Forschungsdaten an der HU

## Was sagen die Forscher?

*"Ich würde meine Daten gerne länger sichern und freue mich, von dieser Möglichkeit zu hören!"*

*"Diese Grundsätze werden während der Ausbildung (Studium) nicht gelehrt."*

*"Ich kenne die Grundätze und halte sie "privat" soweit wie möglich ein, da die Verträge in meiner Karrierestufe jedoch immer befristet und tlw. nicht länger als 2 Jahre sind und übergreifenden Datenbanken bsiher kaum (gar nicht) genutzt werden ist die Praxis der Datenaufbewahrung undefiniert."*

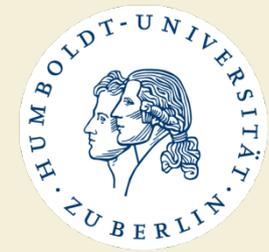
*"Ich weiß nicht, was mit meinen auf dem Server gespeicherten Daten passiert, wenn ich den Lehrstuhl verlasse."*

*"Ich bin gerne bereit, meine Daten zur VErfügung zu stellen, wenn sie "gut behandelt" werden!"*

*"Die Daten werden aufbewahrt, die Grundsätze kannte ich nicht."*

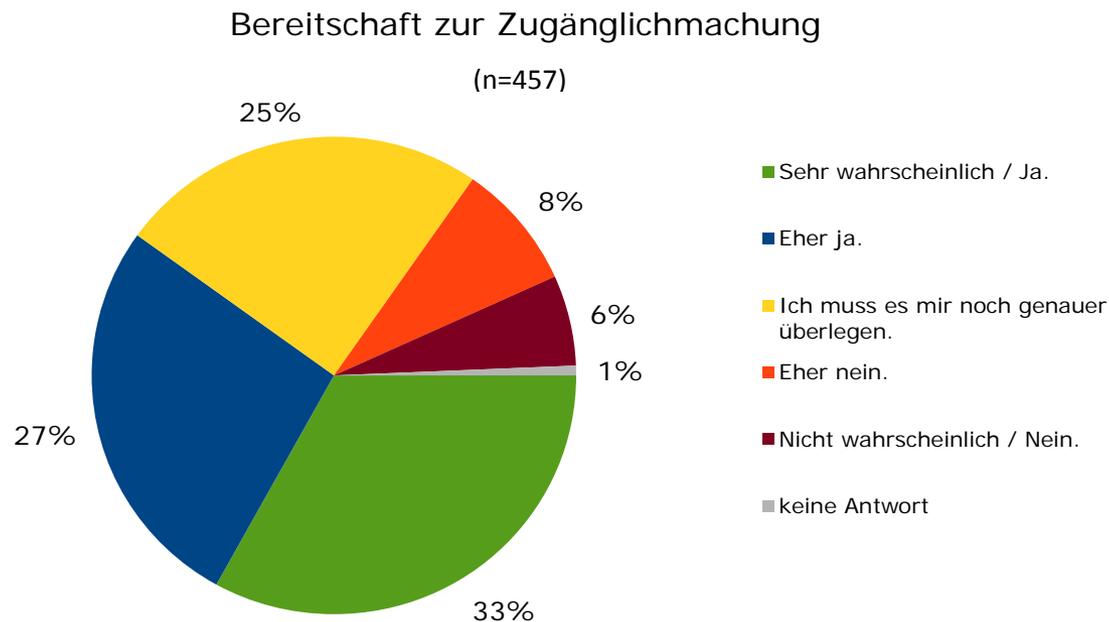


# “It's all about data.”



Umfrage zum Umgang mit digitalen Forschungsdaten an der HU

Wären Sie grundsätzlich bereit, bestimmte Forschungsdaten in einem Datenarchiv abzulegen bzw. zugänglich zu machen?

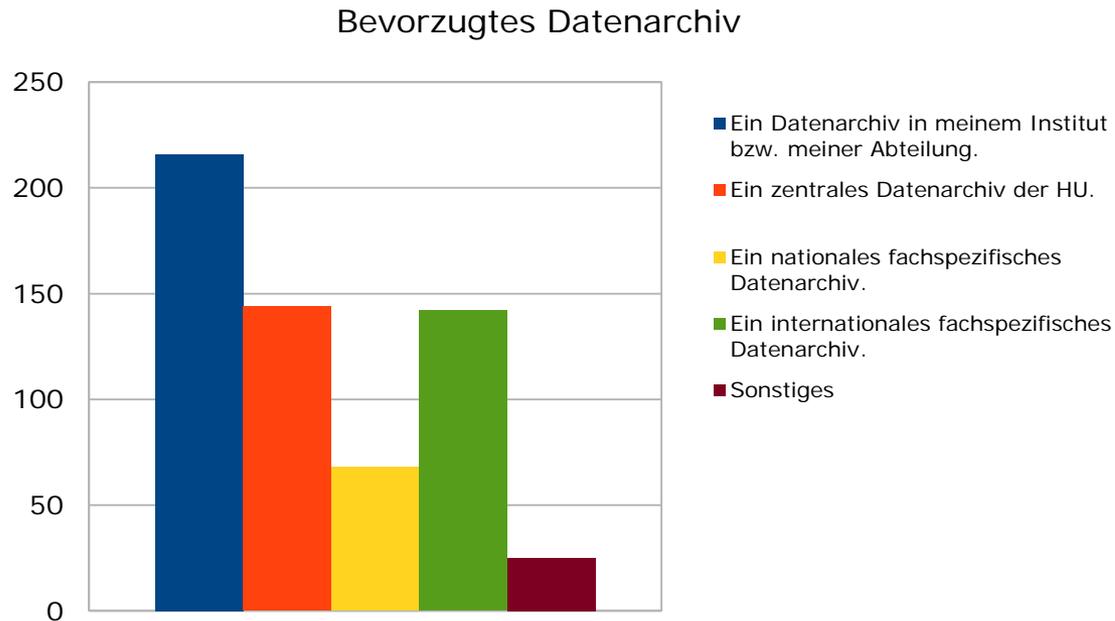


# “It's all about data.”

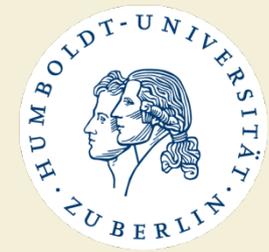


Umfrage zum Umgang mit digitalen Forschungsdaten an der HU

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

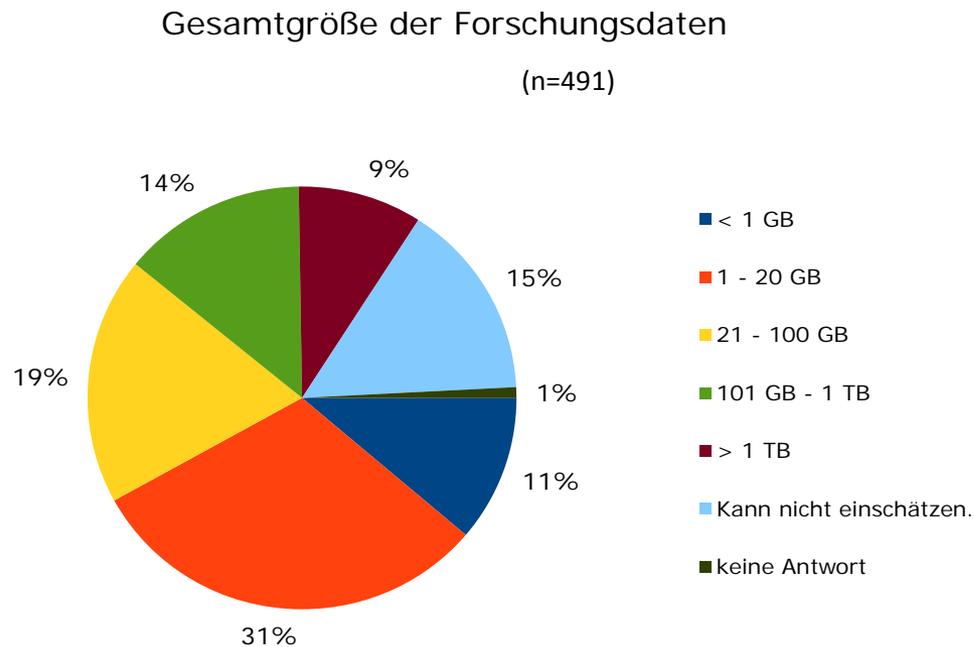


# “It's all about data.”

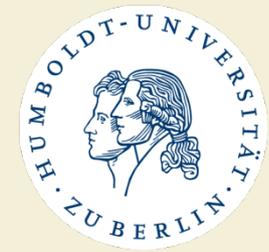


Umfrage zum Umgang mit digitalen Forschungsdaten an der HU

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

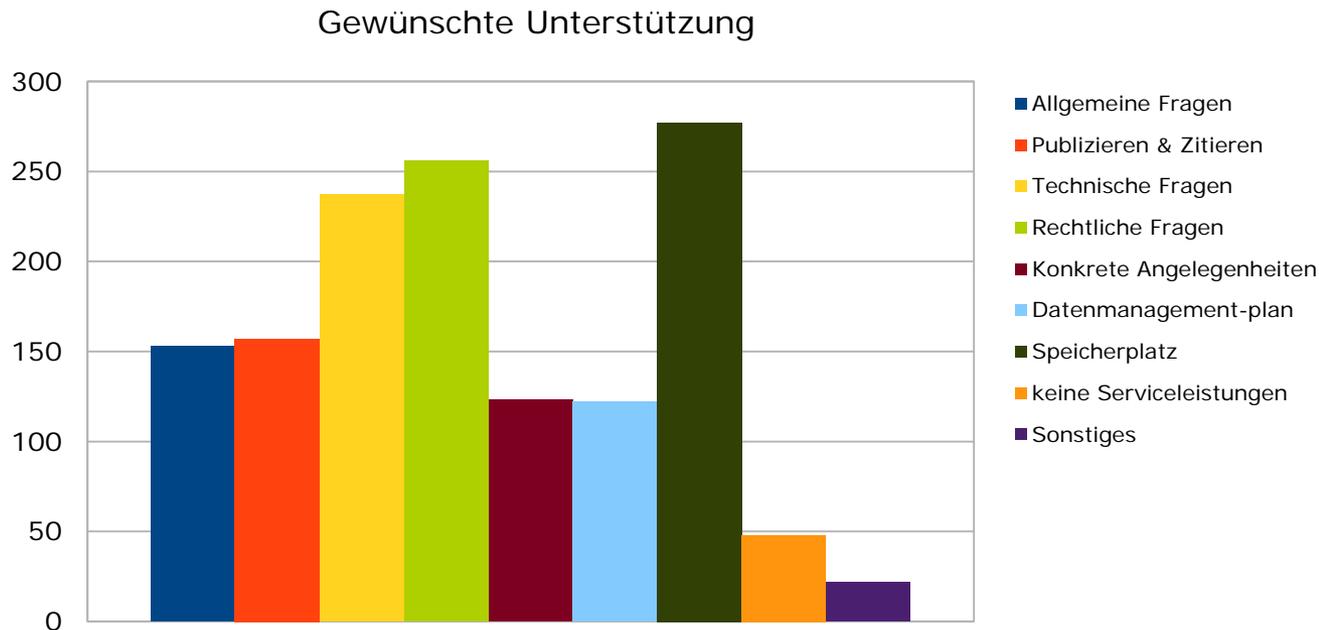


# “It's all about data.”



Umfrage zum Umgang mit digitalen Forschungsdaten an der HU

## Welche Serviceleistungen würden Sie sich von der HU wünschen?



# “It's all about data.”

---



Umfrage zum Umgang mit digitalen Forschungsdaten an der HU

## Weiteres Interesse

*Wären Sie bereit, von Ihren Erfahrungen zum Umgang mit Forschungsdaten und Erwartungen an zentrale Serviceleistungen in einem persönlichen Gespräch zu berichten?*

- Über 70 Zusagen für ein persönliches Interview
- Durchführung von Interviews demnächst

Weitere Kontakte sehr willkommen!

# “It's all about data.”

---



Umfrage zum Umgang mit digitalen Forschungsdaten an der HU

## Erste Schlussfolgerungen für die HU (I)

- Die Humboldt-Universität zu Berlin entwickelt und beschließt eine Policy zum Umgang mit Forschungsdaten.
- Es ist ein Konzept zu entwickeln, das die Forschenden dort abholt, wo sie sich „informationswissenschaftlich“ gerade befinden.
  - Vorhandene Sammlungen/Repositorien sollten in den Forschungsbereichen verbleiben, aber zentral verzeichnet werden.
  - Forschungsbereiche, die eigene Repositorien aufbauen wollen, sollten durch zentrale Einrichtungen auf unterschiedlichste Weise unterstützt werden.
  - Für spezielle Datentypen sollten (unabhängig vom Forschungsgebiet) spezielle Repositorien entwickelt und angeboten werden.
  - Existierende zentrale Angebote sollten ausgebaut und durch entsprechende Services zum Umgang mit Forschungsdaten erweitert werden.

# “It's all about data.”

---



Umfrage zum Umgang mit digitalen Forschungsdaten an der HU

## Erste Schlussfolgerungen für die HU (II)

Verbesserung des zentralen Beratungsangebotes:

- Bei der Umsetzung „Guter wissenschaftlicher Praxis“ und insb. der Dokumentation der Forschungsdaten sollte zielgerichtet in der HU Unterstützung angeboten werden.
- Die rechtliche Beratung zu grundlegenden Fragen des Datenschutzes und des Urheberrechts ist auszubauen.
- Eine zentrale Anlaufstelle für allumfassende persönliche Beratung wird als vorteilhaft gesehen.

# “It's all about data.”

---



Umfrage zum Umgang mit digitalen Forschungsdaten an der HU

## Erste Schlussfolgerungen für die HU (III)

Allgemeine Unterstützung beim Umgang mit Forschungsdaten:

- Vergleichbar mit dem MIT sollten Angebote zu den unterschiedlichen Phasen im Lebenszyklus von Forschungsdaten gemacht werden  
Siehe <http://libraries.mit.edu/guides/subjects/data-management/plans.html>
- Das gegenwärtige Angebot auf den Web-Seiten der HU sollte systematisch gepflegt und aktualisiert werden.  
Siehe: <http://www.cms.hu-berlin.de/ueberblick/projekte/dataman>
- In der HU sollte eine Kapazität aufgebaut werden, damit innerhalb von nationalen und internationalen Aktivitäten entsprechende Beiträge erbracht werden können.

“It's all about data.”

---



## Vielen Dank für Ihre Aufmerksamkeit!

Weitere Diskussion im Anschluss bei informellem Empfang im  
Raum 110 😊

Forschungsdatenmanagement an der HU Berlin  
<http://www.cms.hu-berlin.de/dataman>



21.05.2013

# “It's all about data.”

---



Einige Kommentare und Anregungen für Diskussion

## Wo sind wir heute?

*"Diesen Grundsatz [Gute wissenschaftliche Praxis] beachtet in meiner Gruppe sicherlich niemand. Ich wäre auch der einzige, der Erfahrung mit Haltbarkeit von Datenträgern wie CD etc. [hat]. Einerseits fehlt es an den Organisationsstrukturen dies zu erreichen. Andererseits am technischen Wissen, diese Haltbarkeit mit hinreichender Sicherheit zu erreichen."*

*"Unfortunately there are many incidents of plagiarism in our department and thus I would be quite reluctant sharing data without personal control on who has access on them and why"*

*"Innovations- bzw. Konkurrenzdruck: man hält seine Datenfunde [...] lieber "geheim" und veröffentlicht stattdessen Aufsätze, Bücher etc., in denen die Daten zitiert werden; Ausnahme evt. Gruppenprojekte"*

*"Hauptziel meiner Arbeit sind Publikationen, die nichtveröffentlichten Arbeitsdateien gehen niemanden etwas an und würden auch niemanden interessieren."*

*"Es gibt keine geeigneten Kollaborationssysteme an der HU-Berlin, die zur Mitarbeit von Kollegen anderer Hochschulen animieren."*

*"Manchmal wäre Hilfe in den genannten Bereichen gut, aber ich kann mir auch wieder nicht vorstellen, wie das konkret laufen könnte. Letztendlich quält man sich dann doch wieder allein durch Aufgaben."*



21.05.2013

# “It's all about data.”

---



Einige Kommentare und Anregungen für Diskussion (Fortsetzung)

## Wo wollen wir hin?

*„Zumindest institutsintern würde ich meine statistischen Daten gerne Kolleg/innen und Studierenden zur Verfügung stellen“*

*"Ich würde gerne besser über die Möglichkeiten der Speicherung von Forschungsdaten in Datenarchiven Bescheid wissen und würde diese auch nutzen wollen."*

*"Ich bin für open Access und DATentransparenz, sofern die Urheberschaft mit Kennzeichnung des Projekts gewährleistet ist. Das unterstützt den Austausch, vermeidet Überschneidungen und fördert die Sichtbarkeit meiner wissenschaftlichen Arbeit"*

*"Aber in manchen Faellen sind solche projektgebundenen Infrastrukturen nicht sinnvoll, ein projektuebergreifendes system macht mehr sinn, als in jedem Projekt eine halbgaare Loesung."*

*"Es wäre schon gut, wenn es einen guten cloud Dienst gäbe oder die Netzinfrastruktur ZUVERLÄSSIG über VPN nutzbar wäre."*

*"Nicht nur Beratung sondern praktische Unterstützung"*

*"Einen Ansprechpartner für individuelle Fragen, den man anrufen/aufsuchen kann. Häufig sind meine Frage eine bunte Mischung als allen möglichen hier genannten Dingen."*

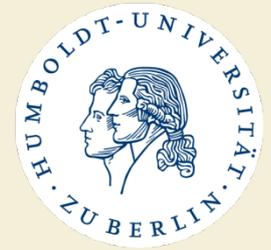
*"Bisher brauchte ich keine Serviceleistungen, aber es wäre sicher gut, sie zu haben."*



21.05.2013

# “It's all about data.”

---



Einige Kommentare und Anregungen für Diskussion (Fortsetzung)

## Was wir zu bedenken haben

*"Der Professor unter dem ich als Doktorand arbeite hat hierbei allerdings das letzte Wort."*

*"Ich habe mich bisher noch wenig mit Datenarchiven befasst, so dass ich erst eine Nutzen/Risiko-Abwägung treffen müsste."*

*"Im Prinzip ja, ich fürchte allerdings, dass ich einige Zeit investieren müsste, um Materialien zu sortieren, bzw. so "aufzuräumen", dass Dritte damit arbeiten könnten oder das Material überhaupt in ein Datenarchiv eingespeist werden kann."*

*"will happily do it once I published my own research using the data."*

*"Der Begriff der Forschungsdaten ist so nicht tragfähig: Wenn alle Texte, mit denen ein Historiker zu tun hat, Forschungsdaten wären, gäbe es enorme urheberrechtliche Probleme mit der Ablage..."*

*"Ich habe Probanden zugesichert, dass niemand außer mir und meiner Arbeitsgruppe Zugriff auf die Daten haben wird."*

*"Quellen unterliegen teils dem Datenschutz/ Urheberrecht. Ich würde den Service nur nutzen, wenn sie vor unbefugtem Zugriff geschützt wären."*

