


Umgang mit Forschungsdaten

Library Lunchtime Learning, Fachhochschule
Nordwestschweiz, Muttenz 25. März 2015

Other Conference Item

Author(s):

Töwe, Matthias 

Publication date:

2015

Permanent link:

<https://doi.org/10.3929/ethz-a-010561400>

Rights / license:

In Copyright - Non-Commercial Use Permitted



Umgang mit Forschungsdaten

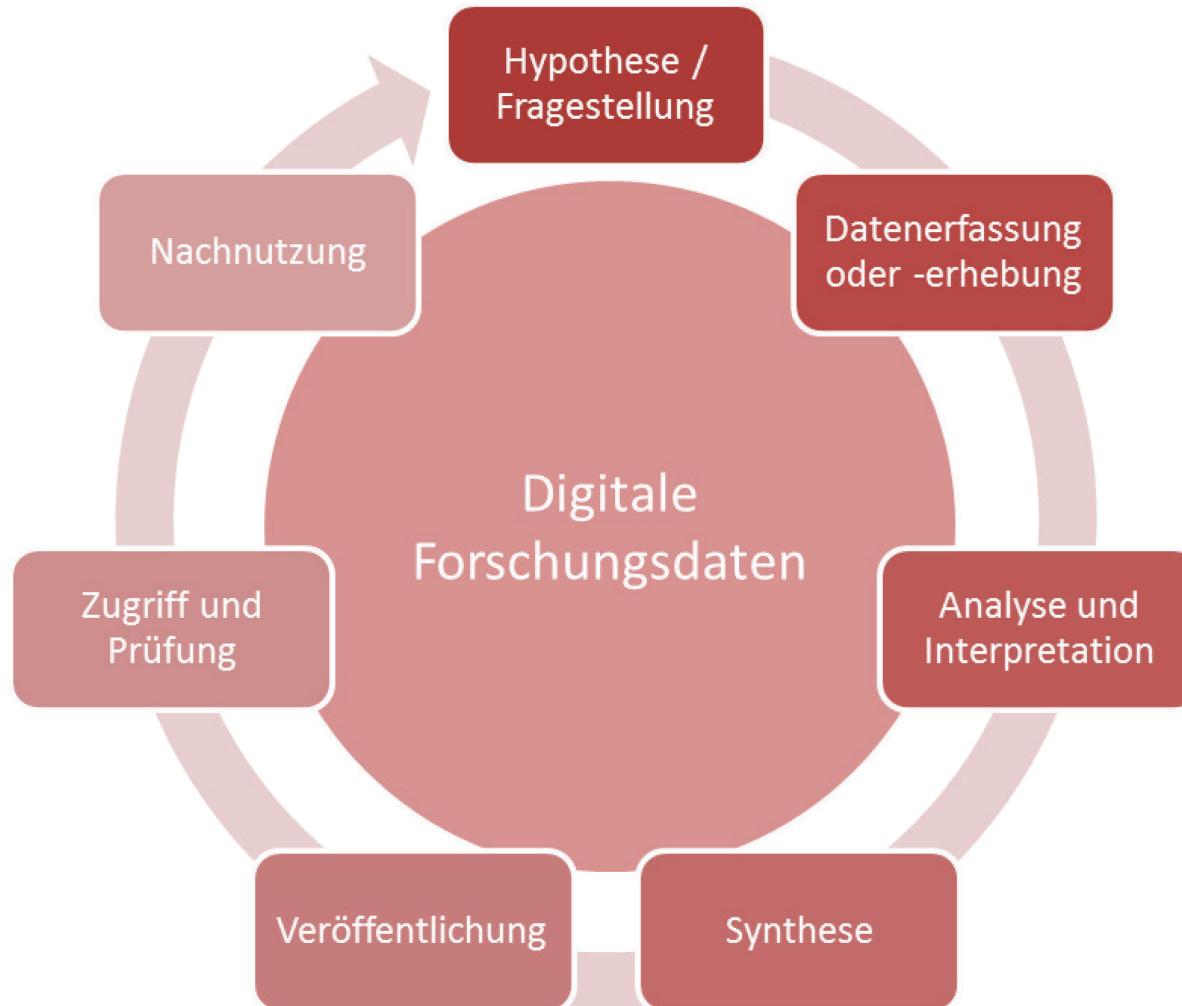
Library Lunchtime Learning, Fachhochschule Nordwestschweiz, Muttenz
25. März 2015

Dr. Matthias Töwe, Leiter Digitaler Datenerhalt, ETH-Bibliothek, ETH Zürich

Überblick

- Lebenszyklus von Forschungsdaten
- Forschungsdatenmanagement
- Publikation von Forschungsdaten / Open Data
- Archivierung von Forschungsdaten
- Rolle der Bibliotheken
- Erfahrungen
- Diskussion

Idealisierter Lebenszyklus von Forschungsdaten

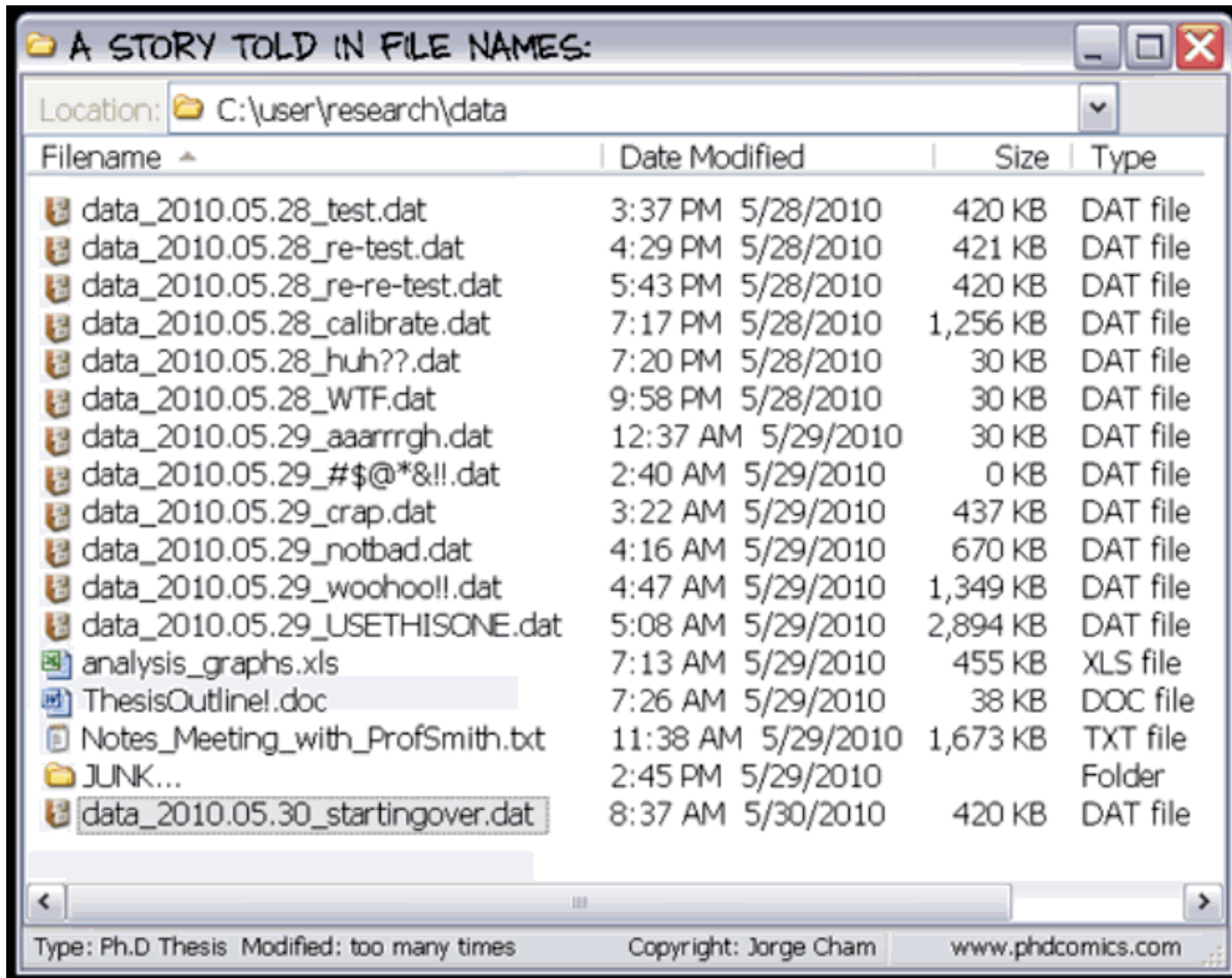


Zwei unterschiedliche Blickwinkel

- Organisieren Sie Ihre Daten während Sie damit arbeiten
- Teilen, erhalten und veröffentlichen Sie Daten für Dritte – und für sich selbst !



Kommt Ihnen das bekannt vor?



"A story told in file names":

Source:

"Piled Higher and Deeper" by Jorge Cham

[www.phdcomics.com
http://www.phdcomics.com/comics/archive.php?comid=1323](http://www.phdcomics.com/comics/archive.php?comid=1323)

Copyright: Jorge Cham
Used with permission.

Forschungsdaten – «Small Data»

- **«Alltagsdaten»**, die jede und jeder Forschende erzeugt
- **Strukturierte Daten** (eher kleinteilig im Sinne von Einzeldateien)
- **Überwiegend für die intellektuelle / manuelle Nutzung**

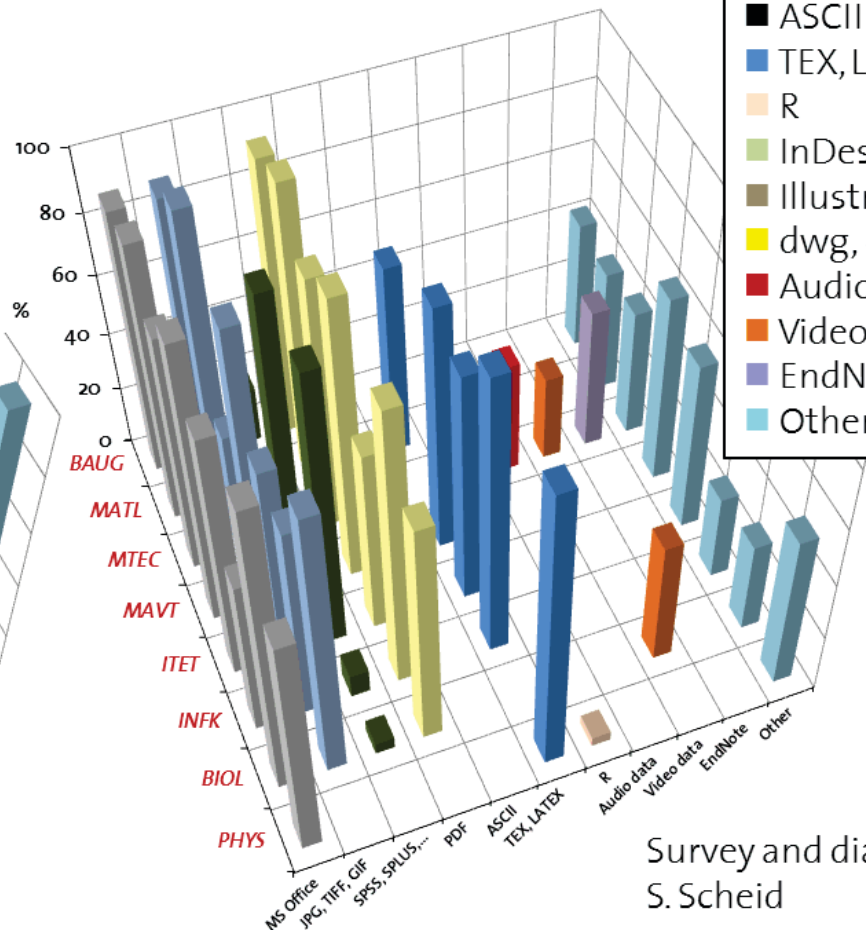
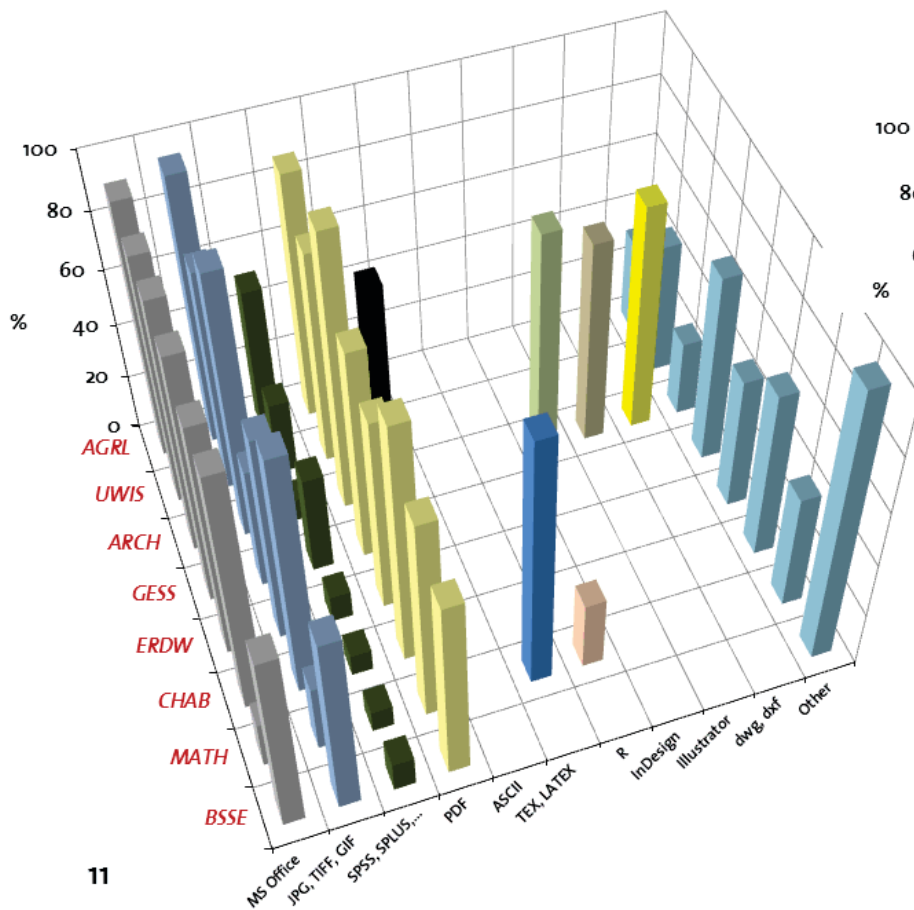
In diesem Fall geht es nicht um:

- **Unstrukturierte Massendaten («Big Data»)**, die regelmässig als Ganzes mit wechselnden Algorithmen prozessiert werden sollen

Genutzte Formate? (Umfrage ETH 2011/12)

2. Part - Which formats do you and your research group work with?

- MS Office
- JPG, TIFF, GIF
- SPSS, SPLUS
- PDF
- ASCII
- TEX, LATEX
- R
- InDesign
- Illustrator
- dwg, dxf
- Audio data
- Video data
- EndNote
- Other



Survey and diagrams:
S. Scheid

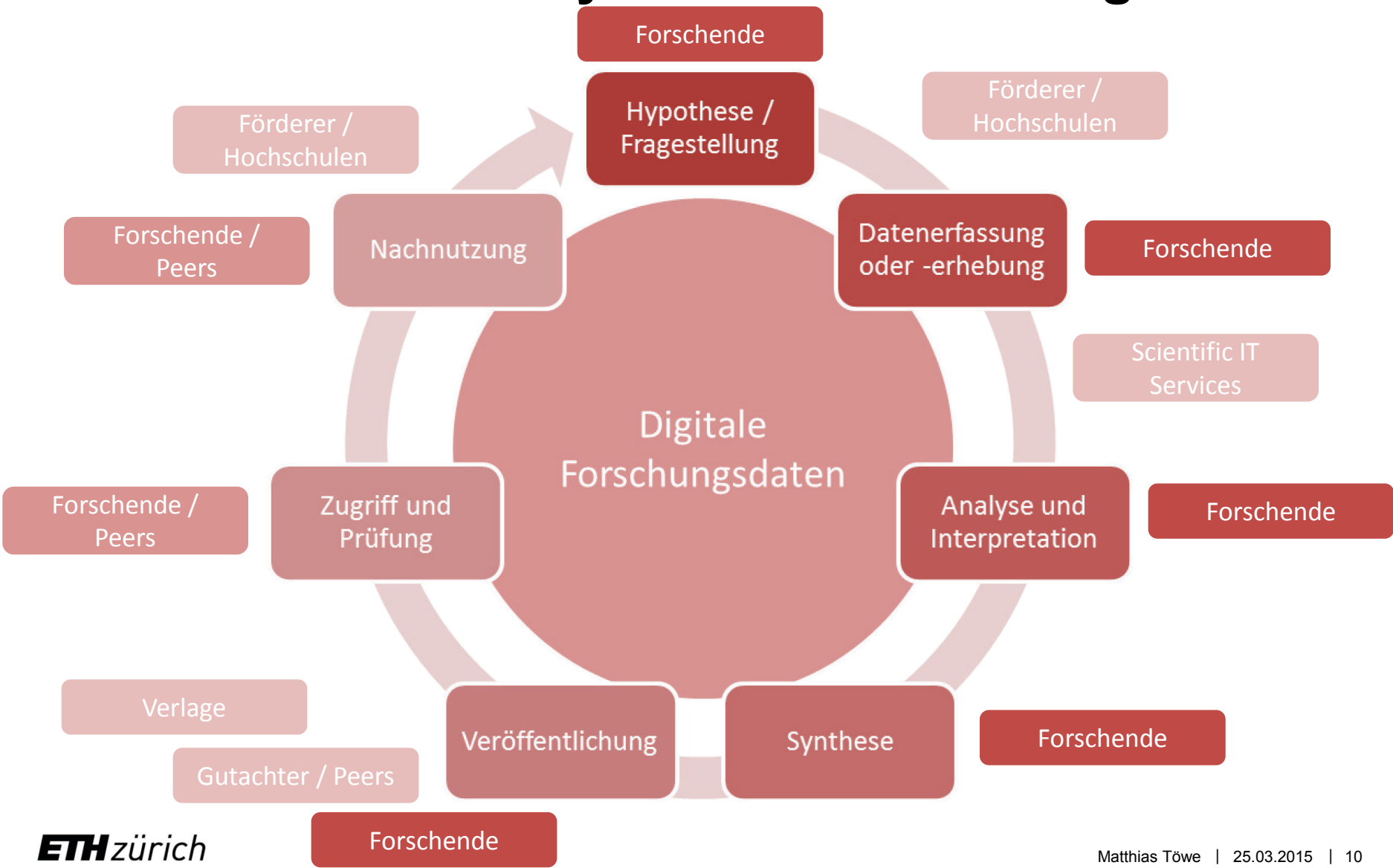
Planung des Umgangs mit Forschungsdaten I

- Fragen zum **Datenmanagement im Projekt**
 - **Welche Daten** werden voraussichtlich erzeugt?
 - **Wie** werden die Daten zunächst **genutzt**?
 - Wie werden lebende Daten **organisiert, dokumentiert und aufbewahrt**?
 - Welche **Regeln** gelten beim Umgang mit diesen Daten?
 - Wie wird die **Aufbewahrung für die Nachprüfbarkeit** gelöst?
- **Fragen bewusst stellen**, vorhandene Tools berücksichtigen
- **Aufwand** budgetieren – und **Zeit** vorsehen
- Viele Antworten vorläufig – müssen **im Projekt aktualisiert** werden

Einige Faustregeln

- Informieren Sie sich über **Standards in Ihrem Fach** und etablieren Sie **zumindest für Ihre Gruppe ein einheitliches Vorgehen**.
- Halten Sie **klare Namenskonventionen** für Ordner ein (z.B. Datentyp, Experiment, Version, Mitglied)
- Begrenzen Sie die **Länge von Ordnernamen** auf 15 bis 20 Zeichen
- Halten Sie eine **Konvention für sprechende Dateinamen** einschliesslich Versionsangabe ein (bis zu 25 Zeichen) und **vermeiden Sie Sonderzeichen**
- Stellen Sie sicher, dass Ihr **Backup arbeitet** und Sie Daten auch wirklich von dort **wiederherstellen** können.
- Nutzen Sie **offene und dokumentierte Formate** → langfristige Nutzbarkeit
- **Vermeiden Sie Verschlüsselungen und Komprimierung** → Nachnutzung

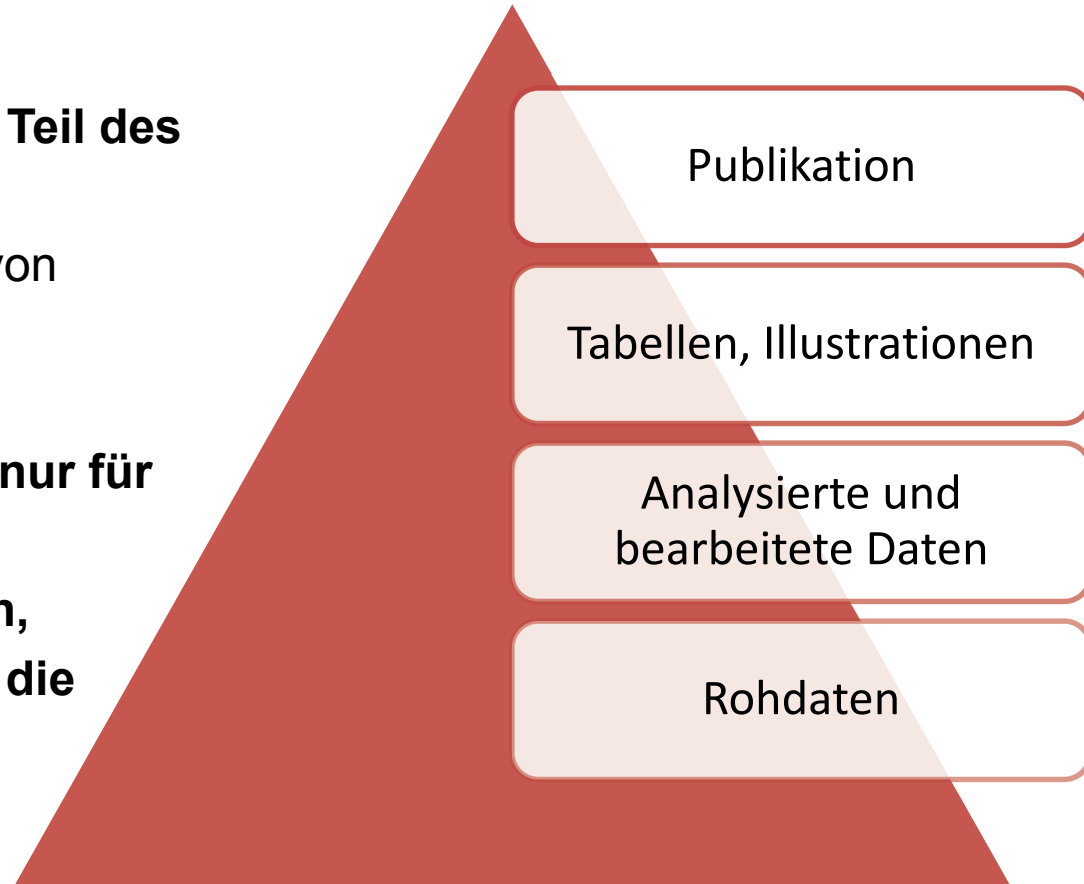
Akteure im Lebenszyklus von Forschungsdaten



Darum geht es langfristig: Daten teilen und dauerhaft nutzbar erhalten

Oft ist nur ein **kleiner Teil des Forschungsoutputs zugänglich** in Form von Publikationen

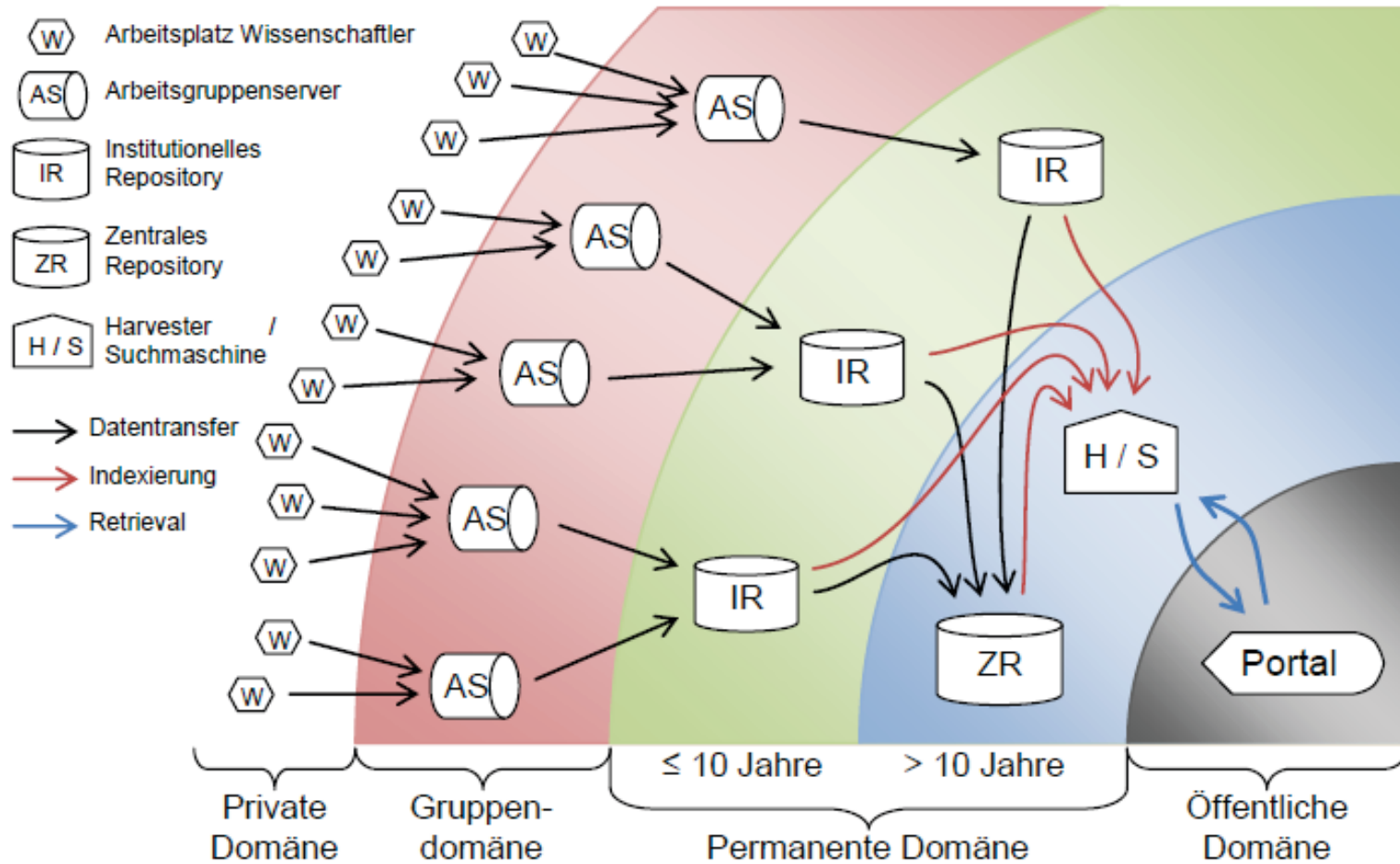
- **Bedenklich nicht nur für Förderer und Steuerzahlerinnen, sondern auch für die Qualität der Forschung**



Planung des Umgangs mit Forschungsdaten II

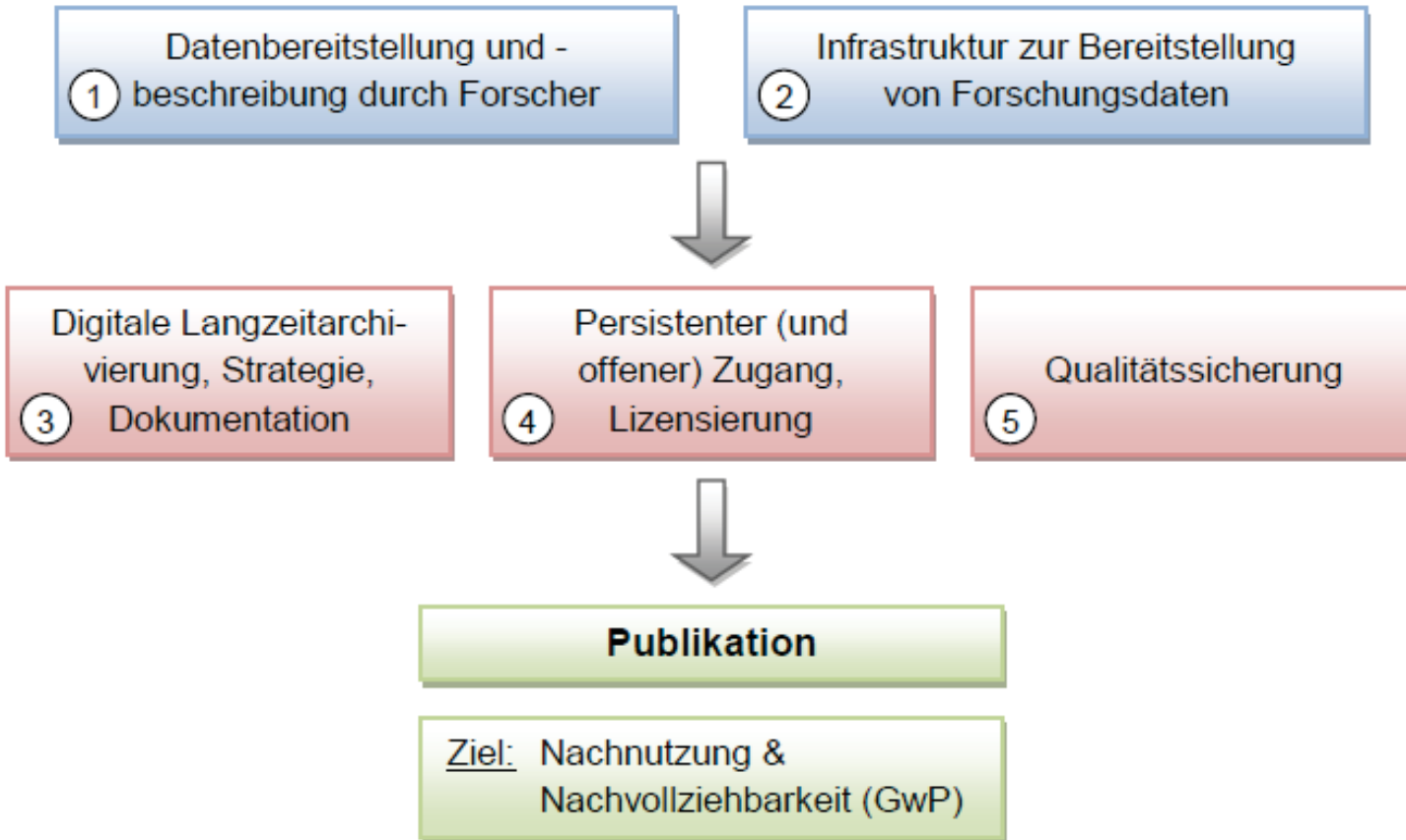
- Fragen zum **Datenmanagement nach dem Projekt**
 - Für welche Daten wird eine (intensive) **Nachnutzung** erwartet?
 - Welche Daten werden **öffentlich** gemacht?
 - Welche Daten erfordern **Anonymisierung** oder Schutzmassnahmen?
 - Welche **Dokumentation** ist nötig für die Nachnutzung?
 - **Welche Repositorien** kommen in Frage (fachliche, institutionelle)?
- **Frühzeitig Vorgaben von Förderern oder Hochschule erfragen**
- **Erfahrungen und Tools der eigenen Community nutzen**
- **Antworten beinhalten immer Annahmen über die Zukunft – einige werden sich als unzutreffend erweisen**

Realistischeres lineares Modell



Modell für eine nationale Struktur in der Chemie, aus: Konzeptstudie Vernetzte Primärdaten-Infrastruktur für den Wissenschaftler-Arbeitsplatz in der Chemie, TIB Hannover, FIZ Chemie Berlin, Universität Paderborn, 2010; S. 100

Was braucht es für die Veröffentlichung?



Aus: Konzeptstudie Vernetzte Primärdaten-Infrastruktur für den Wissenschaftler-Arbeitsplatz in der Chemie, TIB Hannover, FIZ Chemie Berlin, Universität Paderborn, 2010; S. 141

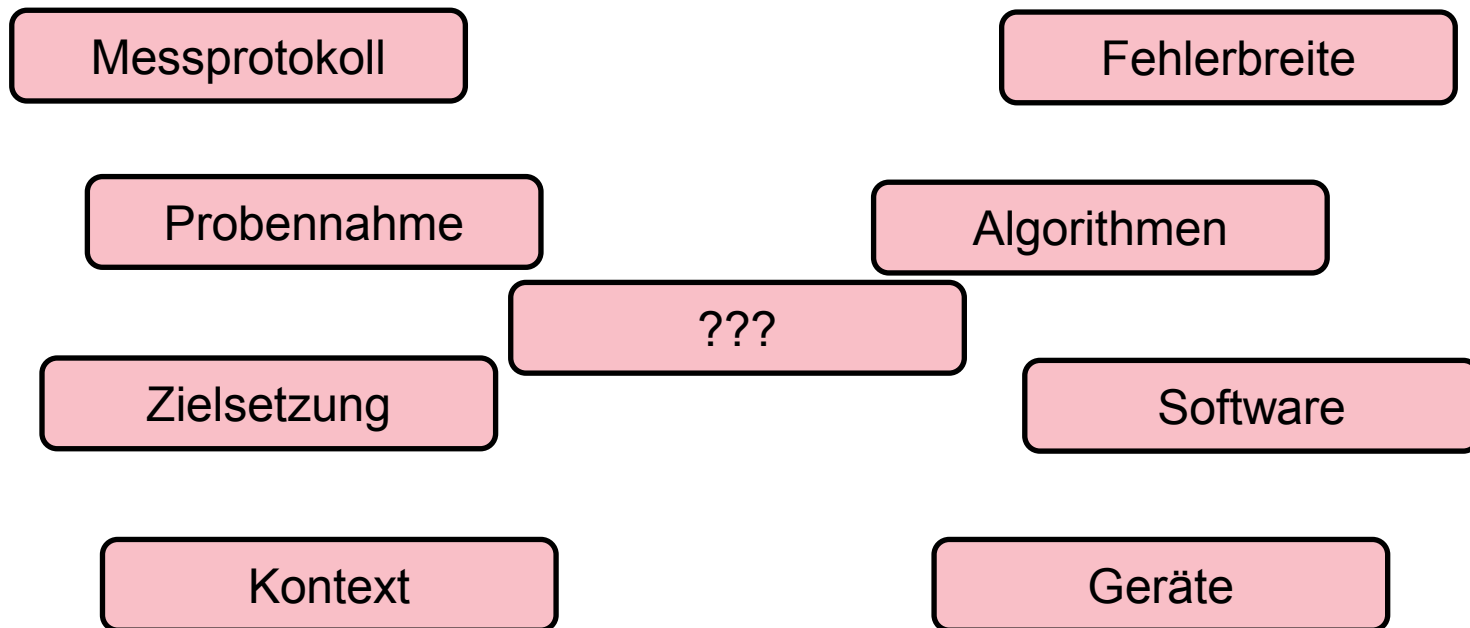
Kommentare von Forschenden zur Nachnutzung von Daten

- *We are struggling to cope with our own data and cannot deal with others' data which requires effort to understand, reconstruct its context, verify it and make sure we have research questions which really fit the data's content.*
- *In our field, **secondary analysis does usually not qualify** for publication in top journals.*
- *We tried once to make our data available in a subject repository, but the **effort to prepare and describe data according to their requirements was too big.***
- *Had we known **how much effort it took** to prepare our data according to the requirements of this journal, **we would have chosen another journal** for publication.*

Was braucht es für die Nachnutzung?

Damit eine andere Person wissenschaftlich seriös mit vorhandenen Daten arbeiten kann, ist eine **umfassende Dokumentation** nötig

→ **Diese wird nur zum Teil durch Publikationen abgedeckt**



Weitergehende Fragen des Vertrauens

Datenproduzenten

- «Kann ich sicher sein, dass ich **für die Nachnutzung meiner Daten ordentlich zitiert** werde?»
- «Werden meine Daten in **wissenschaftlich adäquater Weise** genutzt?»

Datennutzerinnen und -nutzer

- «Kann ich sicher sein, dass die **Daten korrekt verarbeitet** wurden?»
- «Kann ich sicher sein, dass die **Daten zitierbar sind und in n Jahren noch verfügbar?**»

Sharing versus Openness

Zwei unterschiedliche Ebenen des Teilens

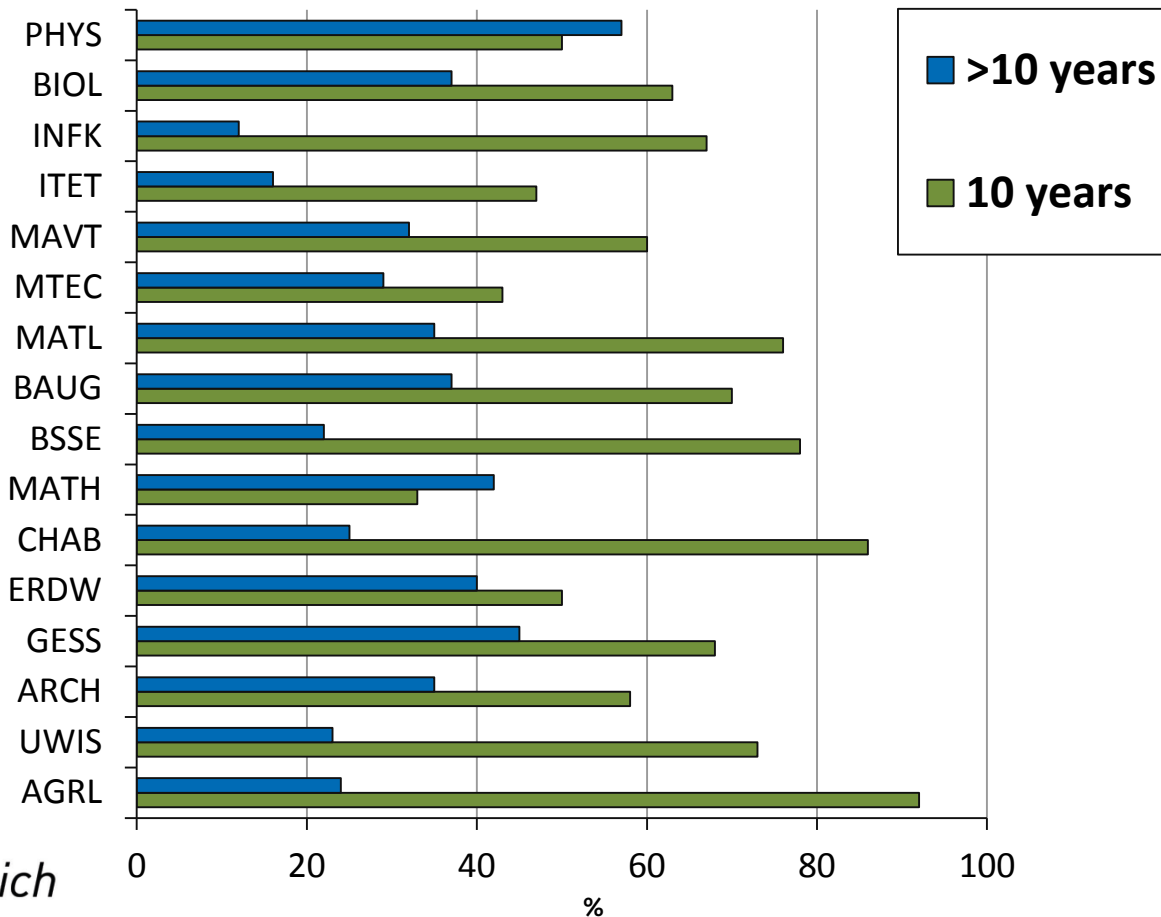
- **Data sharing** mit persönlich bekannten Kolleginnen oder Partnern
 - **Vertrauen** besteht
 - Es gibt ein **gemeinsames Verständnis**, wie Daten verarbeitet und genutzt werden sollen
 - **Für die meisten Gruppen OK**, aber sehr unterschiedlich in der Praxis
- **Daten sind weltweit offen**
 - **Viel mehr Vorbehalte** – und offensichtlich **viel weniger Erfahrung!**

Zwei Seiten einer Medaille?

- Wenn **Daten erhalten** werden, aber für Dritte **nicht offen** sind, sind sie zwar nicht verloren, **nützen aber nur den Produzenten etwas**.
- Wenn Daten **offen zugänglich** gemacht werden, aber **ohne oder mit wenig Kontextdokumentation**, können sie ebenfalls **nur den Produzentinnen nützen** – und selbst die könnten nach einiger Zeit Mühe haben, die Daten zu benutzen.
- Wenn Daten **irgendwo für eine gewisse Zeit zugänglich** gemacht werden, können sie **nicht zitiert oder verlässlich nachgenutzt** werden und die formale Anerkennung (**«Crediting»**) ist **schwierig**.

Gewünschte Aufbewahrungsdauer?

2. Part - How long a period do you or your research group have in mind for storing data?



Survey and diagrams:
S. Scheid

Rolle der grösseren Bibliotheken

- Kompetenzen im **Metadatenmanagement**
- Bereits engagiert
 - als Betreiber von **(erweiterbaren) Repositorien** für Publikationen
 - in Arbeiten zur digitalen **Langzeitarchivierung**
 - bei der Vergabe von **Persistenten Identifikatoren**
 - bei der **Online-Publikation**
- Schnittstellenfunktion:
 - zwischen **Forschung und Publikationswesen**
 - zwischen **Forschung und Verwaltung** (Berichtspflichten usw.)
- **Langfristig angelegte Trägerschaft** und Finanzierung
- ***Das heisst aber nicht, dass sie alles selber machen sollten!***

Ziele der ETH-Bibliothek

- **Nachprüfbarkeit** von Resultaten erleichtern
- Richtlinien für **gute wissenschaftliche Praxis** unterstützen
- Niederschwellige Lösung(en) bieten für die **befristete Aufbewahrung**
- Abstimmung mit / **Abgrenzung zu Informatikdiensten**
- **Datenmanagement** unterstützen
- **Nachnutzung** erleichtern
- **Veröffentlichung** von Forschungsdaten unterstützen
- **DOI-Registrierung** als Teil der Dienstleistung anbieten
- Publizierte Daten **dauerhaft verfügbar halten**
- Wo möglich und sinnvoll: **Aktive Erhaltungsmassnahmen** (Formatmigration) planen
- **Know-how aufbauen** und ETH-Bibliothek als Dienstleisterin in diesem Kontext etablieren

Was braucht es für die Nachnutzung?

Was?

Warum?

Wer?

Data Curation
Datenerhalt

**Intellektuelle
Nachnutzung
ermöglichen**

Datenproduzenten

Content Preservation
Langzeitarchivierung

**Technische
Nutzbarkeit sichern**

**ETH-Bibliothek
(ETH Data Archive)**

**Bitstream
Preservation**
Bit(strom)erhaltung

**Technische Stabilität
gewährleisten**

**Informatikdienste
ETH Zürich**



Angepasst nach Jens Ludwig, Wissgrid

Angebote für ETH-Angehörige

- **Beratung Datenmanagement** (im Aufbau)
- **ETH Data Archive** (Ex Libris Rosetta)
 - **Langzeitarchivierung oder befristete Aufbewahrung** für min. 10 Jahre
 - **Access Rights: Open Access / ETH-intern / individuell / zeitverzögert**
 - **Erhaltungsmassnahmen** (Formatmigration)
 - **Massenprozesse und Einzelaktionen**
 - **DOI-Registrierung** für freigegebene Inhalte via DataCite
 - **Metadaten im Wissensportal der ETH-Bibliothek**
- **docuteam packer** (Open Source)
 - **Viewer und Editor** für lokal erstellte **Dateistrukturen mit Metadaten**
 - **Vorbereitung der Daten für die Übergabe an das ETH Data Archive**

Typische Kundenanfragen

- *«Ich möchte ein **Manuskript für einen Artikel** einreichen. Der Herausgeber verlangt, dass die **Rohdaten in einem Repository** abgelegt werden sollen. Was kann ich tun?»*
- *«Wir archivieren bisher die **Daten zu unseren Dissertationen auf CD-ROM** und möchten eine bessere Lösung finden. Wie kann man das lösen?»*
- *«Wir möchten aus einem Artikel heraus **Forschungsdaten verlinken, die öffentlich zugänglich sein sollen**. Mit unseren eigenen Methoden haben wir sie ausgewertet, jetzt können andere Gruppen mit ihren Methoden weiterarbeiten. Wie machen wir das?»*

Fachhochschulen

- **Teilweise hohe Dezentralisierung als Herausforderung**
- **Abweichende Anforderungen?**
- **Projektantrag «Data Life-Cycle Management» im Programm «Wissenschaftliche Information» der Schweizerischen Hochschulkonferenz**
 - **Partner:**
EPFL, ETHZ, HES-SO (HEG), Unis Basel, Genève (Ltg.), Lausanne, Zürich, SWITCH
 - **Ziel:**
Nutzbare Ergebnisse (Empfehlungen, Leitlinien usw.) für alle und **nutzbare Tools** bzw. Services, vor allem von SWITCH für Angehörige von kleineren Institutionen
 - **Programmbeschreibung und Abstracts der Anträge per Februar 2015:**
<http://www.swissuniversities.ch/de/organisation/projekte-und-programme/suk-p-2-wissensch-information-zugang-verarbeitung-speicherung/>

Erfahrungen

- **Forschungsdatenmanagement ist Teil der Forschung.** Reden Sie mit Fachkollegen im Haus und anderswo auch darüber und **identifizieren Sie Best Practices.** Nehmen Sie Einfluss auf die Diskussion!
- **Schnittstellen** zwischen Forschenden, Bibliothek und IT sind entscheidend: Helfen Sie mit, dass alle **die gleiche Sprache sprechen** oder zumindest merken, wenn sie es nicht tun.
- **Nehmen Sie das Thema Forschungsdaten** individuell und als Institution ernst – aber bleiben Sie pragmatisch: **Etwas konkret Sinnvolles zu tun ist besser**, als nie zum ganz grossen Wurf zu kommen.
- **Wir werden Fehler machen**, falsche Entscheidungen treffen und müssen **daraus lernen** – nichts zu tun ist keine Option.

Mit auf den Weg... (1)

- **Denken Sie frühzeitig** über Ihr Datenmanagement oder das Ihrer Gruppe in einem Projekt nach
- Machen Sie einen groben **Datenmanagementplan** auch wenn Sie dazu noch nicht verpflichtet sein sollten
 - **Was für Daten** erhalten oder erzeugen Sie?
 - **Wie** wollen Sie diese nutzen?
 - **Wie stellen Sie sicher**, dass wichtige Daten erhalten bleiben?
 - Gibt es Daten, die Sie **publizieren** sollten oder möchten?
 - Sollten und dürfen Sie **zusätzliche Mittel für das Datenmanagement** in einem Projektantrag einplanen?

Mit auf den Weg... (2)

- **Nutzen Sie vorhandene Unterstützung**
- Wenn Sie Daten an ein **Online Repository** abliefern sollen:
 - Erkundigen Sie sich nach der **Praxis in Ihrer Gruppe**
 - Sehen Sie auf <http://www.re3data.org> nach einem **passenden fachspezifischen Repository oder Datenarchiv...**
 - ...oder erkundigen Sie sich nach einem **institutionellen oder fachübergreifenden Dienst (z.B. www.zenodo.org)**
- **Aufwändige technische Lösungen sind keine Voraussetzung für echte Fortschritte in der Qualität unseres Datenmanagements.**

Fragen?

Dr. Matthias Töwe
Leitung Digitaler Datenerhalt
ETH-Bibliothek
Rämistrasse 101
8092 Zürich
044 632 60 32

matthias.toewe@library.ethz.ch

<http://www.library.ethz.ch/Digitaler-Datenerhalt>