



# **Langzeitarchivierung von Forschungsdaten - und mehr**

**Bibliothek Information Schweiz: Kongress 2012  
Konstanz, 12.-15. September 2012  
Dr. Matthias Töwe**

# GLIEDERUNG

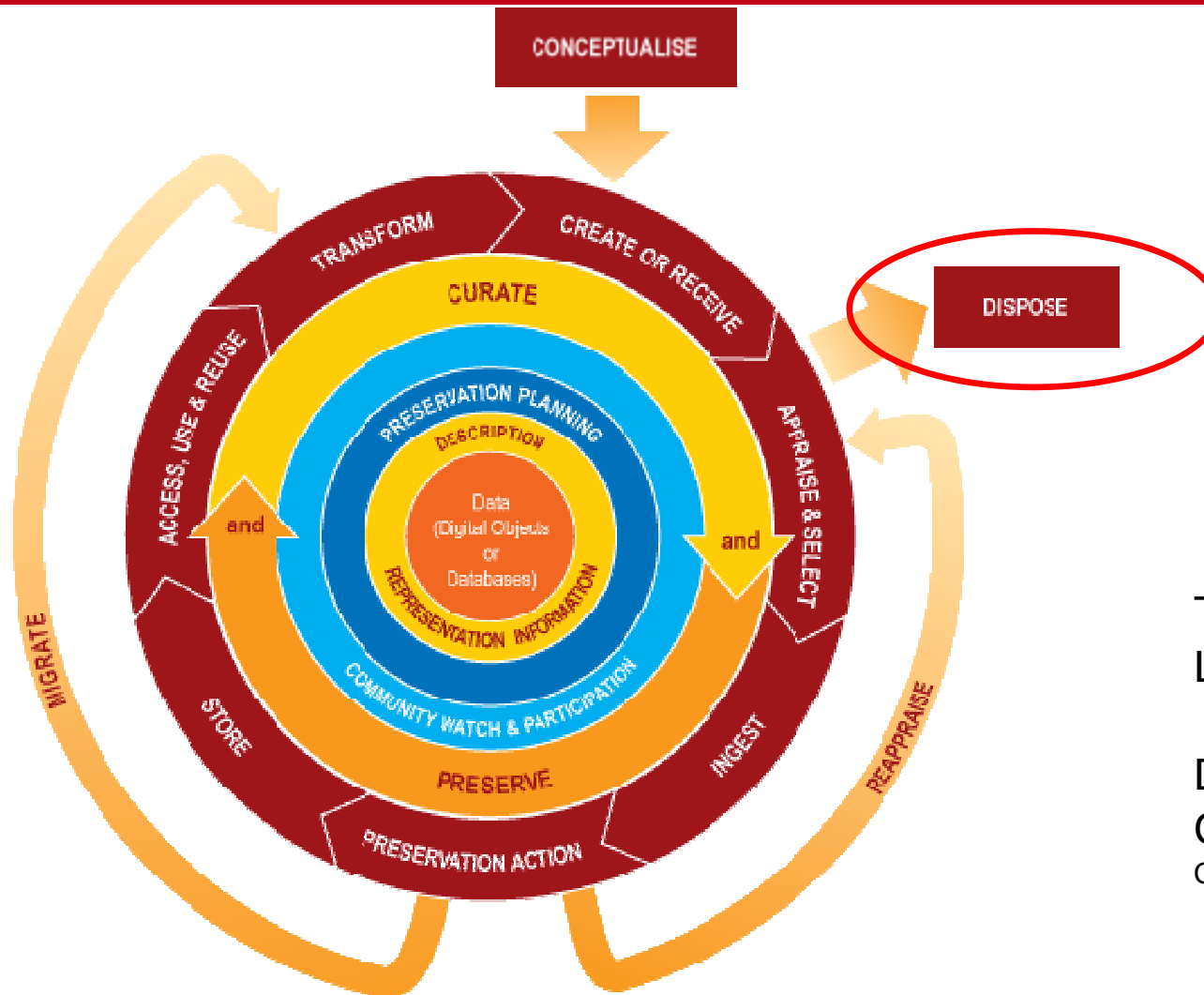
1. Hintergrund
2. Ziele
3. Rolle der Bibliothek
4. Ergebnisse der Umfrage «Umgang mit Forschungsdaten»
5. Lokales Datenmanagement und zentrale Langzeitarchivierung
6. Aktueller Stand

# HINTERGRUND

## Herausforderungen

- **Forschungsprozess stützt sich auf digitale Daten**
- **Gute wissenschaftliche Praxis** verlangt Aufbewahrung von Daten in nutzbarer Form (z.B. Richtlinien ETH Zürich)
- **Förderorganisationen fordern Datenmanagementpläne**
- **Nachnutzung soll gefördert werden**
- Z.T. nicht wiederbeschaffbare **Daten mit dauerhaftem Wert**
- **Veröffentlichte Daten oder referenziertes Zusatzmaterial** müssen zitierbar sein und verfügbar bleiben

# LEBENSZYKLUS



The DCC Curation  
Lifecycle Model

Digital Curation  
Centre, UK  
Copyright 2010

- **Forschende durch Dienstleistung entlasten**
  - **Datenmanagement**  $\leftrightarrow$  **Langzeitarchivierung**
  - Rechenschaft und **Nachprüfbarkeit erleichtern**
  - **Zitierbarkeit** von Daten gewährleisten  
→ DOI-Registrierung bereits in Betrieb  Helping you to find, access, and reuse data
  - Eigene **Nachnutzung**, Zugriff durch Kollegen bis zu echten *Open Data* **unterstützen**
  - **Durchgängige Services** von ETH-Bibl. und Informatikdiensten
- *Erwartung: **Anforderungen** von Förderern und Hochschulen **ans Datenmanagement wachsen***

# ROLLEN IN DER INFRASTRUKTUR

**Übergeordnetes Ziel: lokale Dienste zentral organisieren**

## **ETH-Bibliothek**

- **Informations-, Bibliotheks- und Sammlungsmanagement** für ETH Zürich
- **IT-Services der ETH-Bibliothek** betreiben Bibliotheksapplikationen
- **Projektmandat für die digitale Erhaltung** von der Schulleitung (Universitätsleitung)
- **Schwerpunkt Forschungsdaten**, aber Verantwortlichkeit für alle Daten

## **Informatikdienste der ETH Zürich**

- **IT Infrastrukturmanagement**
- **IT Dienstanbieter** einschliesslich der virtuellen Umgebung für Rosetta
- **Speichermanagement**

# BEDÜRFNISSE DER FORSCHENDEN\*

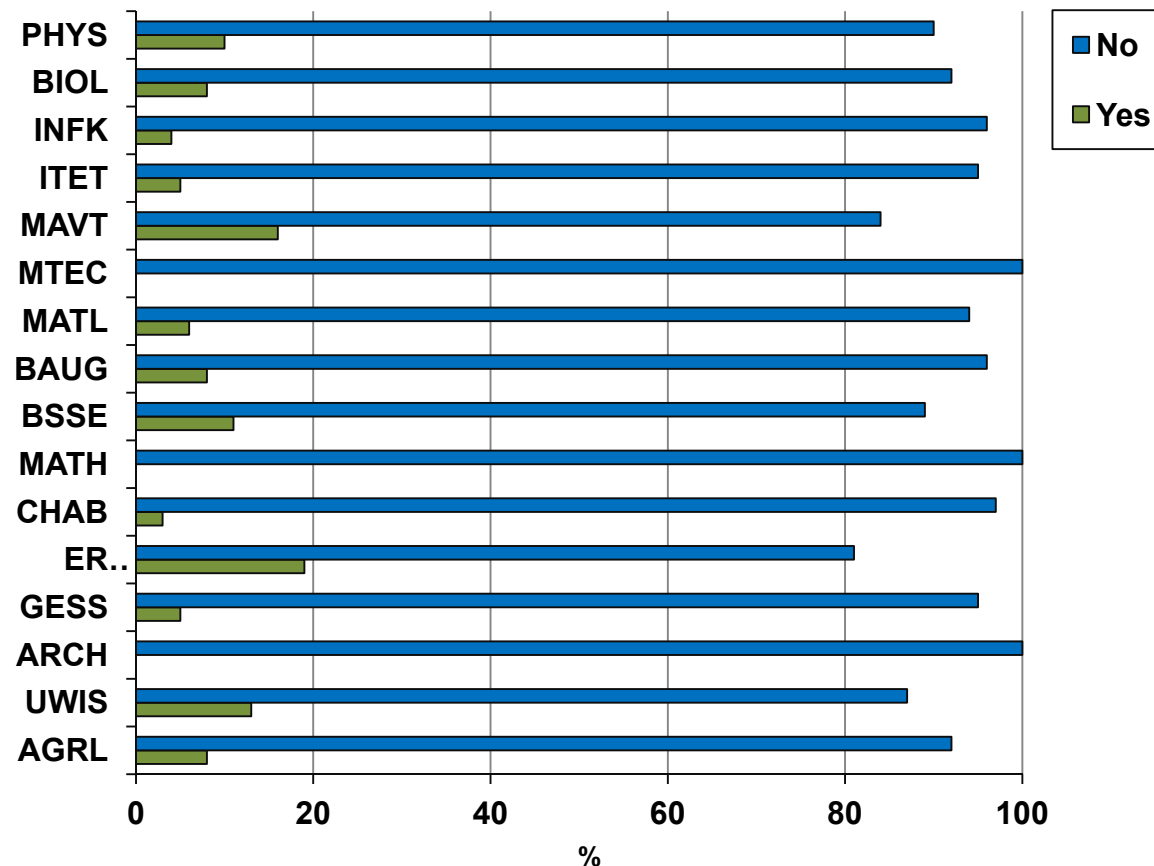
## Forschende...

- müssen viele Daten für **begrenzte Zeit** aufbewahren (z.B. 10 bis 12 Jahre)
- möchten **Daten, die publiziert werden, dauerhaft referenzieren und archivieren**
- möchten Daten vor dem Ingest **umstrukturieren, auswählen, dokumentieren**
- möchten **Metadaten editieren und Daten hinzufügen**
- möchten **Kontrolle** darüber behalten, wer auf ihre Daten zugreift
- sind **interessiert an Unterstützung bei der Erhaltung und Qualitätskontrolle** (Checklisten, Rückmeldungen zu den Metadaten...)
- wollen **keinen zusätzlichen Arbeitsaufwand** ohne Mehrwert

*\*(Gemäss Umfrageantworten von 80% der Forschungsgruppen und Diskussion mit Pilotpartnern)*

# UMFRAGE: POLICIES DER FORSCHUNGSGRUPPEN

**1. Part** - Does your research group have a written 'Data management policy' on handling its own or external research data, or similar documents describing how to handle research data?

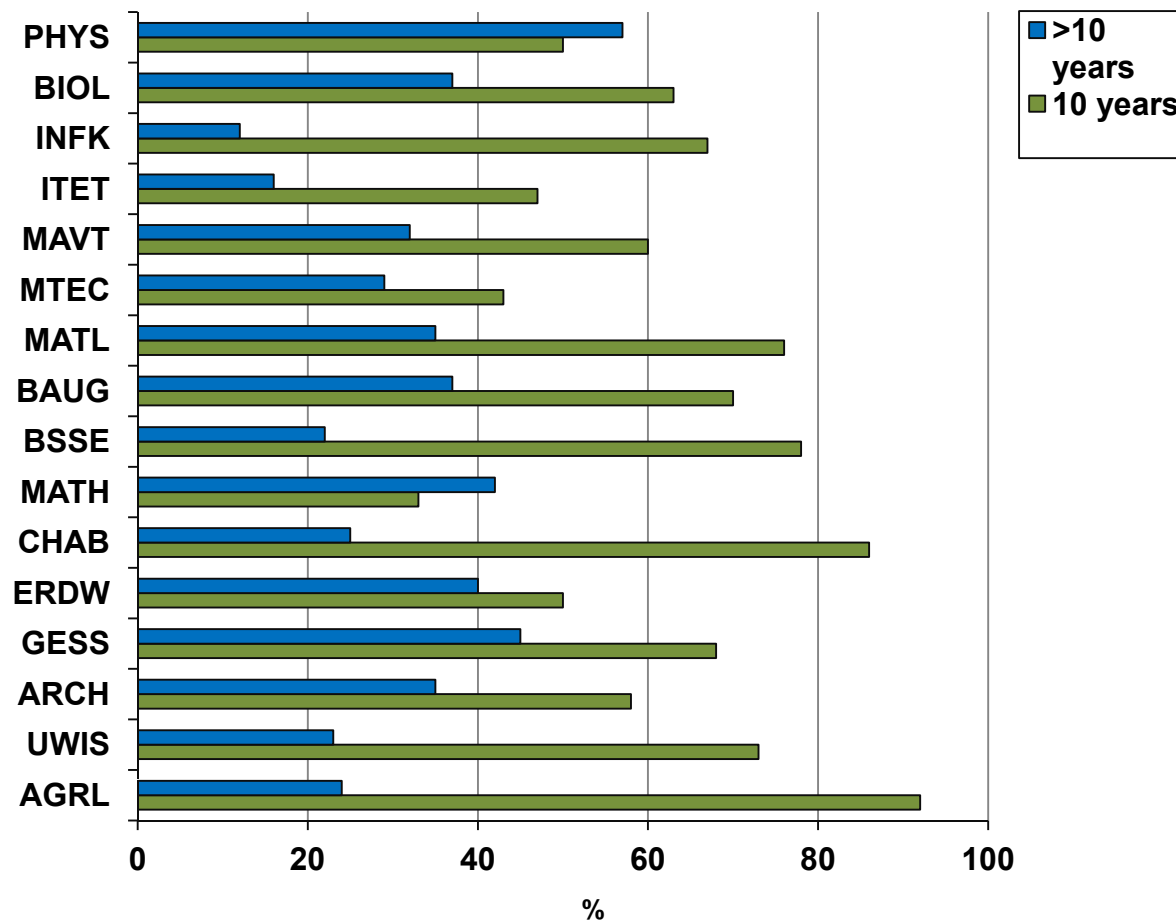


S.Scheid



# UMFRAGE: AUFBEWAHRUNGSDAUER

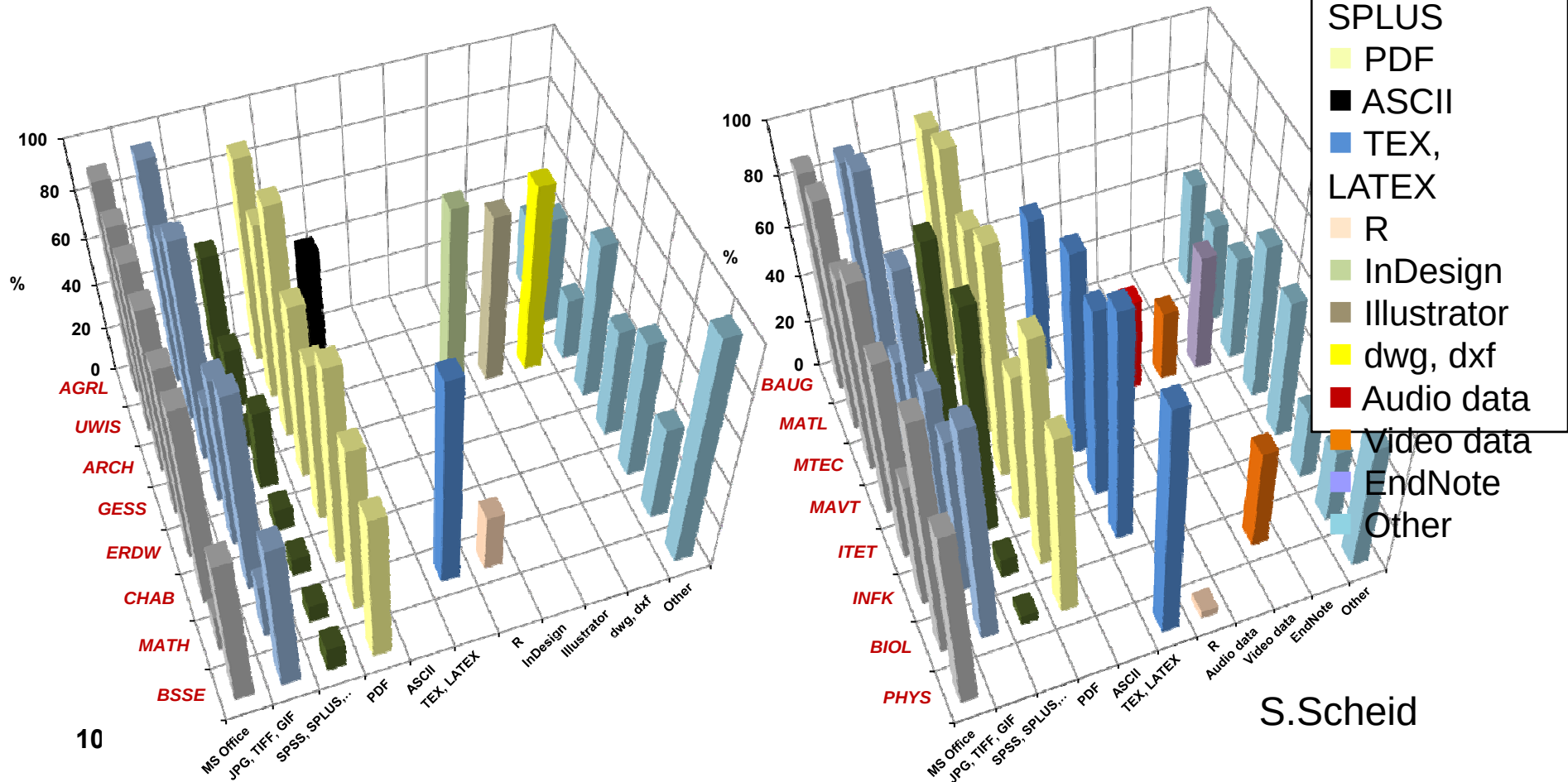
2. *Part* - How long a period do you or your research group have in mind for storing data?



S.Scheid

# UMFRAGE: DATEIFORMATE / SOFTWARE

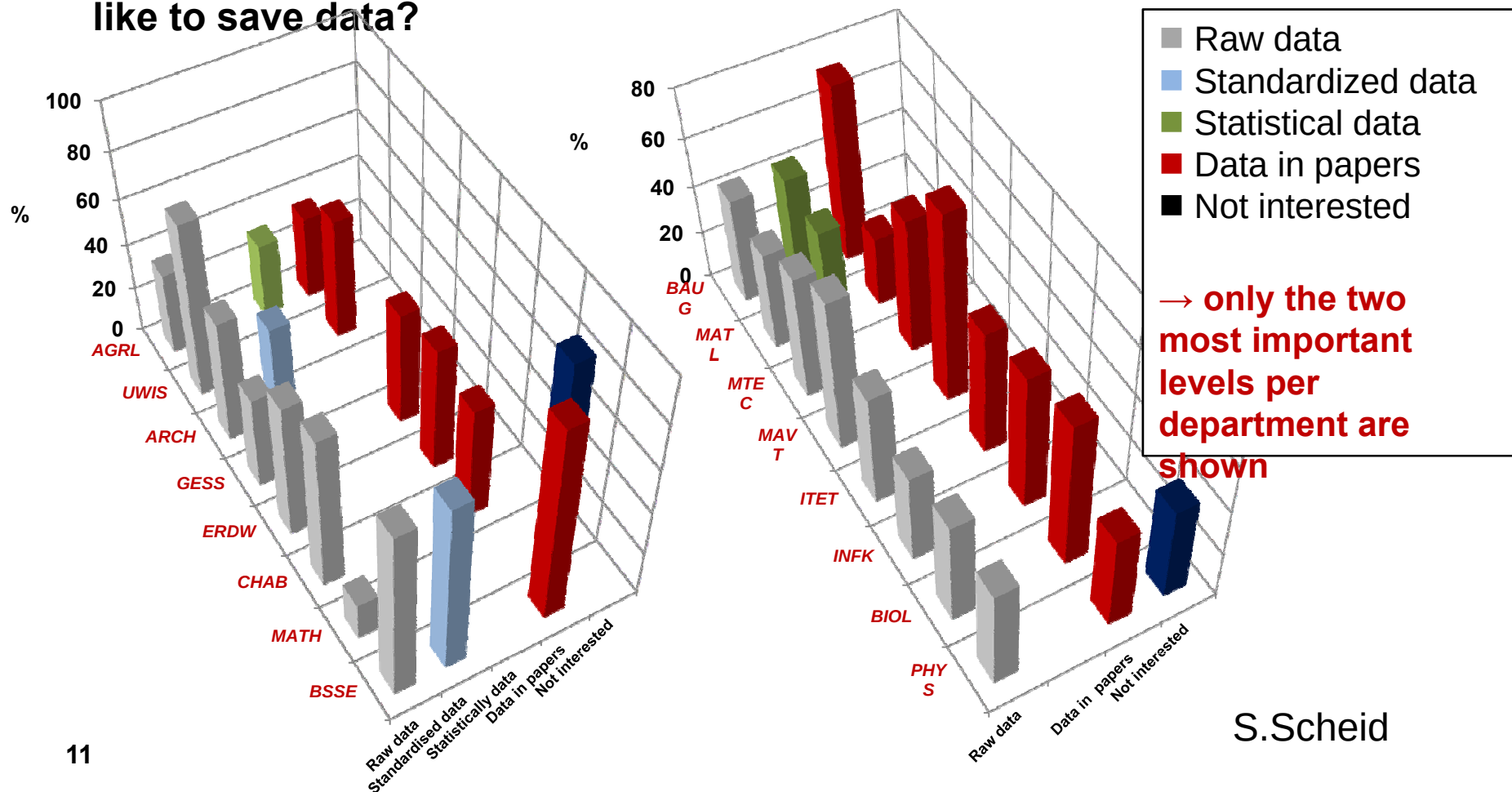
## 2. Part - Which formats do you and your research group work with



S.Scheid

# UMFRAGE: STATUS DER DATEN

**2. Part** - If the ETH-Bibliothek were to provide you with a database for storing your research data, at which level would you or your research group like to save data?



S.Scheid

# FUNKTIONALE EBENEN

**Was?**

**Warum?**

**Wer?**

**Data Curation**

**Ensure intellectual  
re-usability**

**Data Producers**

**Content  
Preservation**

**Ensure technical  
re-usability**

**ETH-Bibliothek  
(Rosetta)**

**Bitstream  
Preservation**

**Ensure technical  
stability**

**IT-Services  
ETH Zurich**

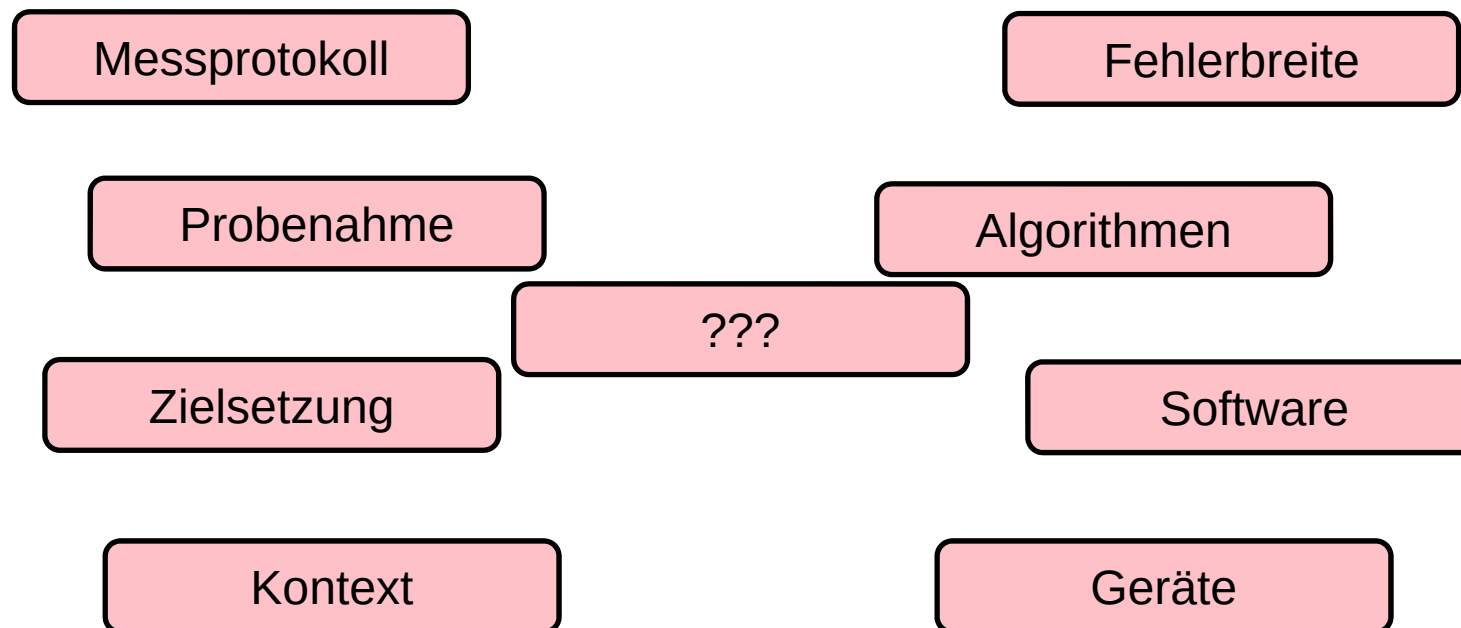


Adapted after Jens Ludwig, Wissgrid

# INHALTLICHE SICHT

**Umfassende Dokumentation notwendig:**

- **Beispiel: apparative Messung einer physischen Probe**  
→ *idealerweise* durch Publikationen abgedeckt, *real* nur bedingt



# EIGENSCHAFTEN DER MATERIALIEN

## Forschungsdaten

- Kontext nur teilweise bekannt
- Vielfältige Formate
- Oft keine Standardformate
- Erzeugung durch Dritte, nur lang-sam und begrenzt zu beeinflussen
- Wenige formalisierte Metadaten
- Rechte häufig unklar (Eigentum, Nutzung, Veröffentlichung)
- Öffentlicher Zugriff nicht automatisch erwünscht
- Vorbehalte bei Datenproduzenten möglich

## Bibliotheksobjekte

- Kontext meist bekannt
- „Wenige“ Formate
- Meist Standardformate
- Erzeugung kann schnell beeinflusst werden
- Formalisierte Metadaten
- Rechtesituation meist klar
- Öffentliche Bereitstellung ist meist das Ziel
- In-house: Ein gewisses Vertrauen wird vorausgesetzt – oder verfügt

# UNTERSCHIEDE ZWISCHEN DEN DATENTYPEN?

**Was?**

**Forschungsdaten Bibliotheksobjekte**

**Data Curation**

**Comprehensive  
documentation by  
producers required**

**Full control of  
metadata and  
context**

**Content  
Preservation**

**More and less  
common formats**

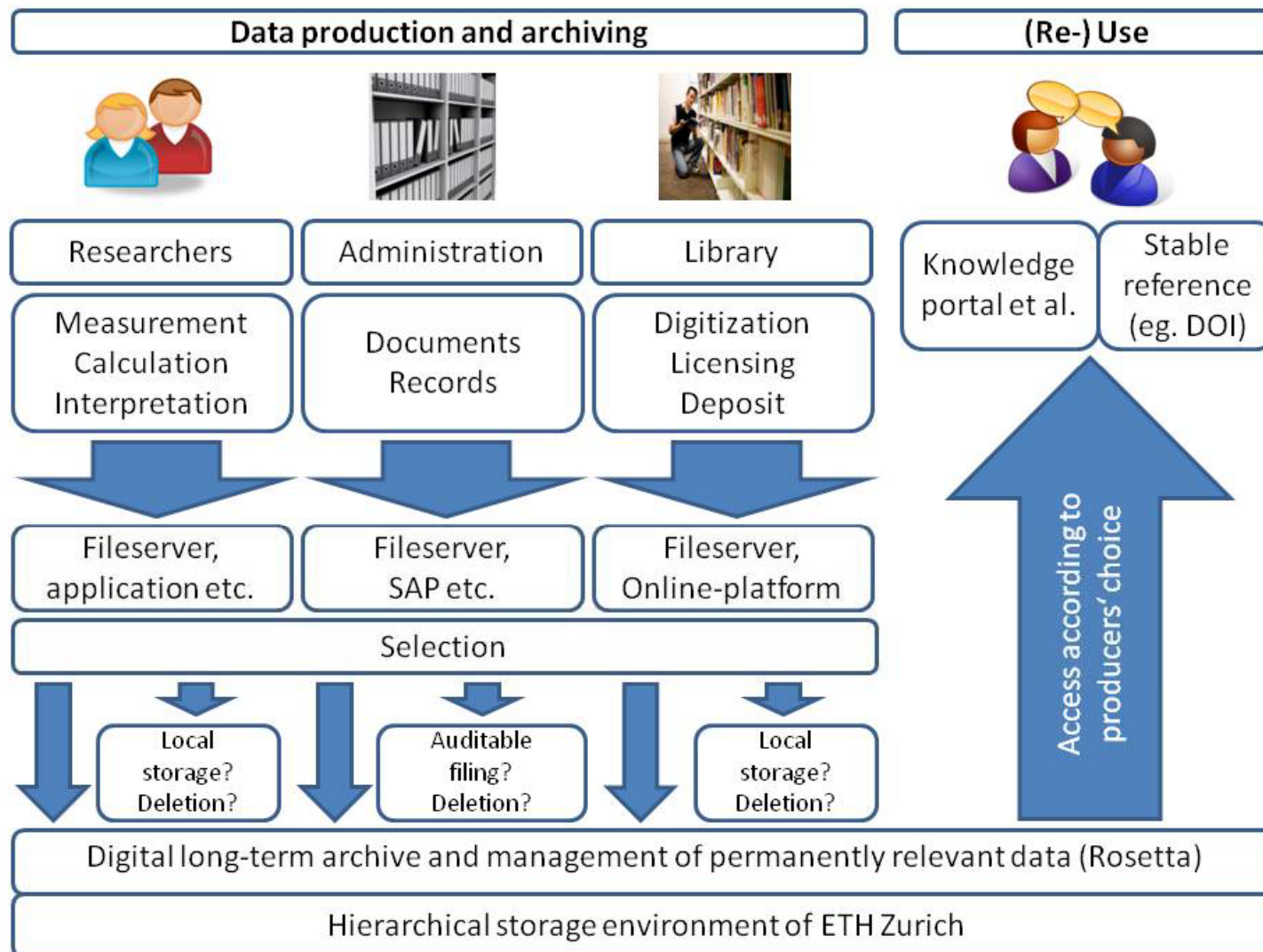
**Mainly standard  
formats**

**Same preservation procedures apply**

**Bitstream  
Preservation**

**„Any object is just bits“**

# VISION: ROSETTA ALS GEMEINSAME BASIS





# DATENMANAGEMENT UND OAIS?

- **Mit Pilotpartnern erarbeitete Anforderungen** betreffen kaum Funktionen innerhalb des OAIS-Rahmens
- Dagegen hohe **Flexibilität erforderlich im Pre-Ingest** oder davor
- **Auswirkungen auf die Rolle des LZA-Systems Rosetta:**
  - **Wenig sinnvoll, Komplexität für die Langzeitarchivierung durch neue Funktionen weiter zu erhöhen**
  - Sofern vorhanden, **Daten aus vorgelagerten Anwendungen übernehmen**
  - Bei Bedarf **Flexibilität im lokalen Datenmanagement** erreichen, nicht in zentraler Anwendung

# LOKALES DATENMANAGEMENT?

- **SIP Package Handler** («Docupack» (Java))
- **Viewer und Editor** für Strukturierung und Metadatenerfassung
- **Daten und Metadaten** beliebig lange **lokal** auf eigenem Speicher jeder Forschungsgruppe
- **Erzeugt** bei Anstoss der Archivierung **SIP (Submission Information Package)** für den Ingest in Rosetta
- **Herkunft aus dem archivischen Bereich** kein Zufall: Anforderungen sehr ähnlich

# «UNDER CONSTRUCTION»

Forschen  
de

Bibliothek

Strukturierung  
und Metadaten-  
erfassung,  
DOI-Erzeugung

Zeitpunkt X:  
Auswahl für  
Archivierung  
und «Submit»  
inkl. Access  
Rights im Archiv

Ingest, DOI-  
Registrierun  
g

Rosetta

Lieferung  
via DOI,  
wenn  
Rechte ok

Gewählter  
Speicher (lokaler  
PC, Share);  
Berechtigungen

über Dateisystem

M. Töwe

**ETH**-Bibliothek

Eidgenössische Technische Hochschule Zürich  
Swiss Federal Institute of Technology Zurich

# ERWARTUNGEN AN DIE DIENSTLEISTUNG

- **Gewichtung von Services gegenüber Softwarebereitstellung**
- Bedarfsorientierte, **individuelle Konfiguration vs. Handhabbarkeit und Automatisierung** der Gesamtlösung
- Nicht **einfach ein weiteres Tool** zur Verfügung stellen
- Forschende erwarten **echte Unterstützung und Services**, die ihre Arbeit erleichtern...
- ... und eine möglichst **nahtlose Integration** in ihre Arbeitsabläufe
- ... sowie Lösungen nicht nur für Forschungsdaten, sondern auch für **administrative Unterlagen**

# STRATEGISCHE FRAGEN

- Wie sieht das Profil der Dienstleistung «Datenerhalt» aus?
- **Bibliothek verändert ihre Rolle** noch stärker von der Informationsversorgerin hin zur Dienstanbieterin → bewusste Gestaltung nötig
- Positionierung der **Bibliothek in** einem stark **IT-lastigen Arbeitsfeld**
- **Kosten** der aufwändigen Dienstleistung müssen **gedeckt** werden – gleichzeitig soll das Angebot **keine Kostenbarriere** aufweisen
- **Zusammenarbeit** von Forschenden, Informatik Support Leitern, Informatikdiensten und Bibliothek, wenn es auch um Geld geht?

# Globale Herausforderungen

- **Formatanalyse, -konvertierung und Ergebnisvergleich**
  - Verfahren noch **nicht ausgereift**
  - **Know-how wenig verbreitet**
- **Daten und ihr Management in dauernder Entwicklung**
- **Langzeitarchivierung als Prozess** ohne einmalige und dauerhafte «Lösung», sondern als **Gegenstand der Forschung**
- **Wie geeignetes Personal aufbauen?**  
Forschungsnah, bibliothekarisch und archivarisch, IT-affin...
- **Kultur und Methoden der qualifizierten Nachnutzung?**

# BISHERIGE UND AKTUELLE SCHRITTE

- ✓ **DOI-Registrierung** durch ETH Zürich als Mitglied von DataCite
- ✓ **Umfrage bei** allen Forschungsgruppen (Profs.) der ETH Zürich
- ✓ Identifizierung von **Pilotpartnern; Ermittlung ihrer Anforderungen**
- ✓ Prüfung und Aktualisierung **des Inventars von Daten der Bibliothek**
- ✓ **Submission application** für die ETH E-Collection
- ✓ **«Gap analysis»** und Entscheidung über Entwicklungen

- **Entwicklung und Testen** der Erweiterungen für Rosetta
- **Entwicklung und Testen** des lokalen Datenmanagements

**Sofern erfolgreich:**

- **Ausdehnung der Abdeckung** auf weitere Gruppen
- Übergang vom Projekt zur **produktiven Dienstleistung**

# VIELEN DANK !

## Fragen ?

Dr. Matthias Töwe  
Leitung Digitaler Datenerhalt  
ETH-Bibliothek  
Rämistrasse 101  
8092 Zürich  
044 632 60 32  
[matthias.toewe@library.ethz.ch](mailto:matthias.toewe@library.ethz.ch)  
<http://www.library.ethz.ch>