



Data Life Cycle Management – national sowie an der UZH

Fachtagung anlässlich des 100-Jahr-Jubiläums der Zentralbibliothek Zürich:

«Die Bibliothek vernetzt. Infrastrukturen für Forschungsdaten in den Geisteswissenschaften»

9. Februar 2017

**Christian Fuhrer
Universität Zürich
Hauptbibliothek
Open Access
Strickhofstrasse 39
8049 Zürich**

www.hbz.uzh.ch, www.oai.uzh.ch



Dieses Werk ist lizenziert unter einer [Creative Commons Namensnennung 4.0 International Lizenz](https://creativecommons.org/licenses/by/4.0/).



Abstract

Forschungsdatenmanagement (Data Life-Cycle Management, DLCM) bedeutet umfassenden und nachhaltigen Umgang mit digitalen Forschungsdaten. Daten sollen verlässlich erfasst, gespeichert, kategorisiert, versioniert, geteilt, archiviert und publiziert werden unter Berücksichtigung fachspezifischer Besonderheiten. In den meisten Fachgebieten bilden digitale Daten inzwischen die Grundlage, gehören zur Arbeitsweise und stellen einen wichtigen Teil des Forschungs-Outputs dar. Geldgeber wie der Schweizerische Nationalfonds und die Europäische Kommission verlangen vermehrt Datenmanagement-Pläne bei der Einreichung eines Forschungsgesuchs.

Im gross angelegten DLCM-Projekt des SUK P-2 Programms arbeiten unter der Leitung der Universität Genf acht Partnerinstitutionen der Schweiz zusammen, darunter Hauptbibliothek und Zentrale Informatik der UZH. Ziel ist der national koordinierte Aufbau von Dienstleistungen zum DLCM. Leitlinien, Hilfestellung zu Datenmanagement-Plänen, Schulungsunterlagen und weitere Informationen sollen auf einer zentralen Website zur Verfügung stehen.

Im lokalen Data-Pilotprojekt erarbeiten Zentrale Informatik und Hauptbibliothek der UZH zusammen mit der Zentralbibliothek Zürich Lösungsansätze für die Implementierung von DLCM an der UZH. Dabei spielen Kontakte zu repräsentativen Forschungsgruppen, die Klärung konkreter Anforderungen, die Definition von Workflows sowie Tests von Plattformen aus der Community eine grosse Rolle.

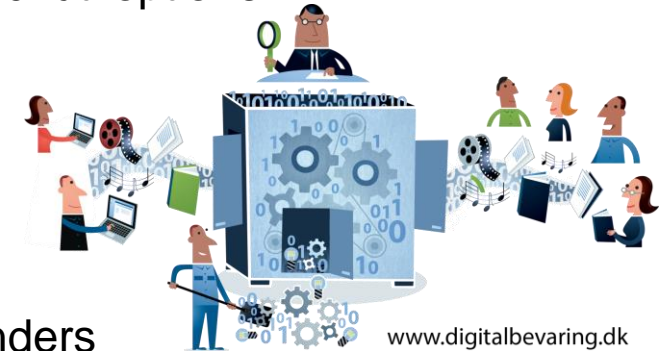


Data Life Cycle Management, Forschungsdatenmanagement: Worum geht es?

Um etwas, das eigentlich alle Forschenden betrifft, je nach Disziplin etwas anders gewichtet:

Forschungsdaten, Quellenmaterialien, Versionen, Annotationen, Inhalt von Datenbanken, ...

- erfassen, bearbeiten, verlässlich speichern mit Backup- und Sicherheits/Datenschutzoptionen
- finden – nicht verlieren, auch bei Personalwechsel
- kontrolliert teilen – mit der Forschungsgruppe, in Zusammenarbeiten
- ausgewählte Datensätze publizieren, langfristig und sicher archivieren
- rechtliche Fragen klären
- Unterstützung erhalten bezüglich Anforderungen von Förderorganisationen/Funders



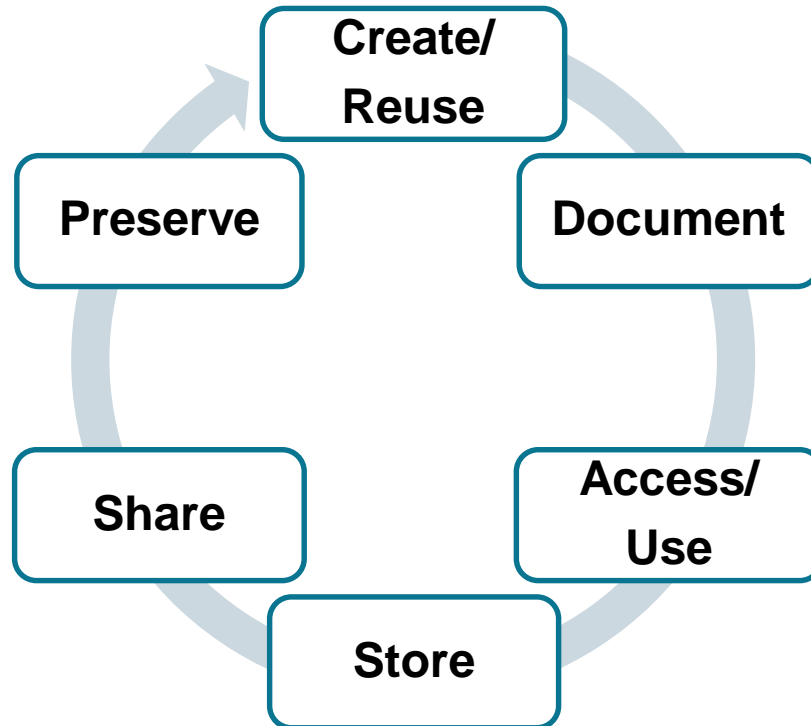
Diese Praxis gestaltet gleichzeitig die Wissenschaft allgemein effizienter und nachvollziehbarer!

Trotzdem wenden sie viele Forschenden zurzeit nicht oder zu wenig an:

- weil ein gewisser Aufwand besteht und sie im Moment keine Zeit dazu haben
- weil eine geeignete niederschwellige Infrastruktur fehlt
- weil Anreize und Reputationsgewinn nicht vorhanden sind



Data Life Cycle Management, Forschungsdatenmanagement



- **Data Management Planning**
- Creating/Reusing data
- **Documenting** data
- Accessing / using data
- Storage and backup
- Sharing data
- Publishing **Open Data**
- **Preserving** data

Forschungsdatenmanagement = Grundlage zu **Open Data** und **Open Science**

Grosse Themen in Europa: Wissenschaftsminister, EU Horizon 2020, LERU, European University Association...



Forschungsdatenmanagement beim SNF

**Datenmanagement-Pläne werden zunehmend bei der
Gesuchseinreichung verlangt...**

“Da der bessere Zugang zu Forschungsdaten zur Qualität der
Forschung beiträgt, neue Forschungsfragen ermöglicht und generell
zu Effizienzgewinnen führt,

**wird der SNF soweit sinnvoll einen Plan für das
Datenmanagement als Bestandteil eines Gesuches verlangen,**

sich an Kosten für die Datenablage beteiligen sowie die Entwicklung
von Infrastrukturen im Bereich Datenmanagement verfolgen und
gezielt unterstützen.”

<http://www.snf.ch/de/derSnf/publikationen/Seiten/mehrjahresprogramm-17-20.aspx>





Data Life Cycle Management national

Bestehende oder entstehende Services für Fachgebiete und Teilaspekte

- **Swiss Center for Expertise in the Social Sciences (FORS)**: stellt für die entsprechenden Fachgebiete in der Schweiz einen Service zur Veröffentlichung und Nutzung von Forschungsdaten zur Verfügung.
- **Swiss Art Research Infrastructure (SARI)**: nationale und internationale Forschungsinfrastruktur zur Vernetzung digitaler Bildquellen und Metadaten im Bereich der Kunstgeschichte, Architektur, Archäologie, Film-, Geschichts-, Literatur- und Religionswissenschaft und anderen bildwissenschaftlich arbeitenden Disziplinen.
- **Nationale Infrastruktur für Editionen (NIE/INE)**: Im Rahmen des SUK P-2 Projektes soll im Sinne einer nationalen Infrastruktur für Editionen (NIE - INE) eine Arbeits- und Publikationsplattform geschaffen werden, die grundsätzlich allen wissenschaftlichen Editionsprojekten in der Schweiz offen steht.
- **Data and Service Center for the Humanities (DaSCH)**:
 - Preservation of research data in the humanities and their long-term data curation.
 - Ensuring permanent access to research data in order to make it available for further research and thus facilitating the reuse of existing research data in future research.
 - Providing services for researchers to assist them with the data life cycle management.
 - Encouraging the digital networking of databases created in Switzerland or in other countries.
 - Collaboration and networking with other institutions on digital literacy.

–



Data Life Cycle Management national: SUK P-2 DLCM



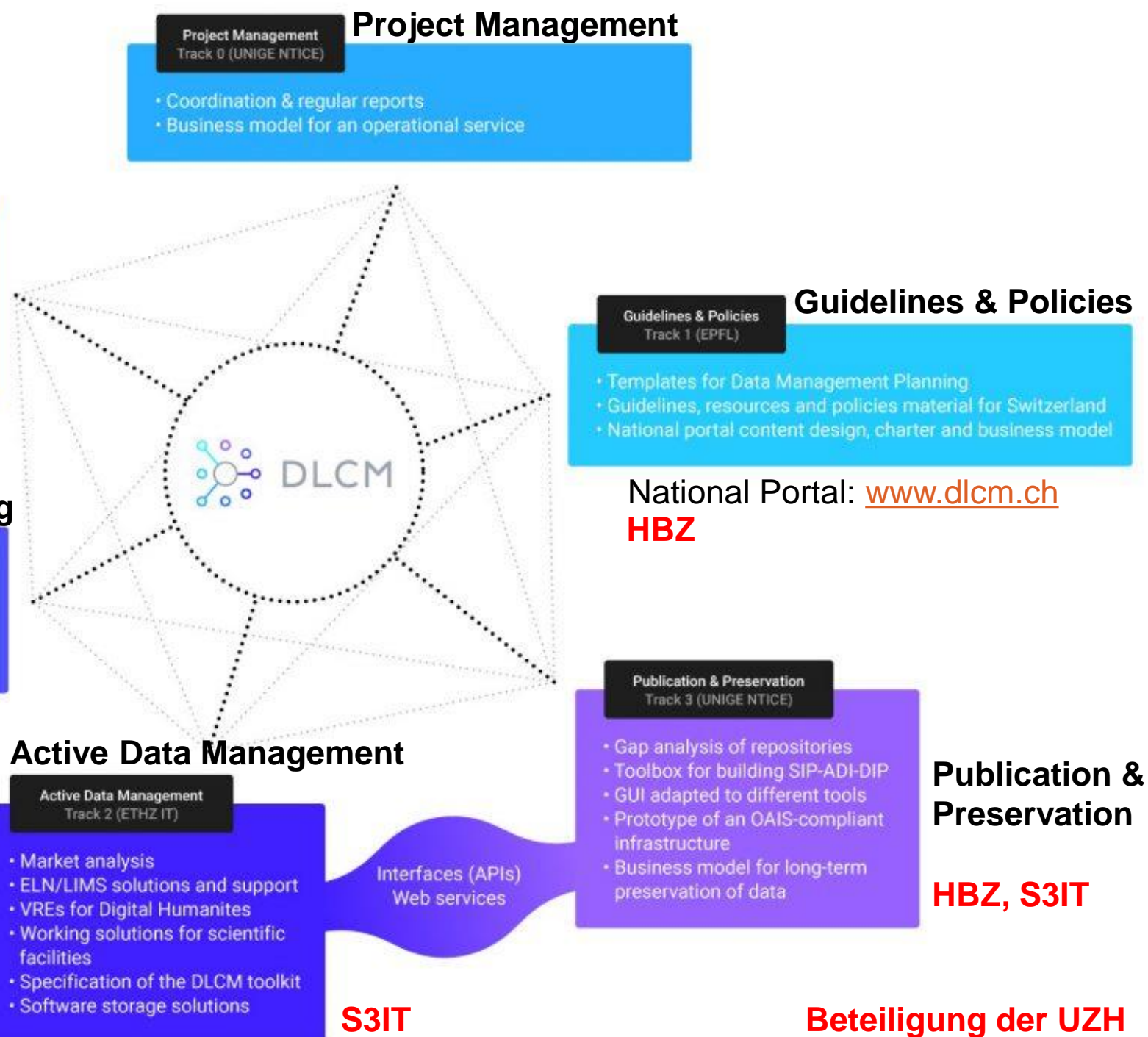
Genereller entstehender Service für alle Fachgebiete

SUK P-2 DLCM: Das Projekt Data Life-Cycle Management des SUK P-2 Programms

(Wissenschaftliche Information: Zugang, Verarbeitung und Speicherung) von swissuniversities und SHK befasst sich mit allen Aspekten von Forschungsdatenmanagement – von Erfassung, Speicherung, Aufbereitung bis zum Teilen, Publizieren und langfristigen Archivieren. Es enthält Anwendungsfälle aus Geisteswissenschaften, Naturwissenschaften und Medizin.

- Aktive forschungsnahe Daten, virtuelle Forschungsumgebung --- stabile Daten zur Archivierung und/oder Veröffentlichung ---- Beratung, Schulung, Bewusstseinsbildung
- Laufzeit: 1.8.2015 – 31.7.2018, Gesamtbudget ca. CHF 10 Mio, ca. 50% Eigenmittel
- Kooperationspartner: Universitäten Genf (Lead), Lausanne, Basel, Zürich, ETH, EPFL, FH HEG & HES-SO, Switch
- Oft zwei Teilnehmer (Science IT und Bibliothek) pro Institution. Beispiel UZH:
S3IT (Zentrale Informatik)
Hauptbibliothek: André Hoffmann, Martin Brändle, Christian Fuhrer
- Austausch mit anderen (fachspezifischen) Initiativen, z.B. DaSCH
- Wie die nationalen Services konkret aussehen werden, ist noch in Entwicklung
- Nachhaltigkeit noch offen, Geschäftsmodell in Entwicklung

SUK P-2 DLCM: Tracks, Ziele



Outreach & Dissemination

HBZ

HBZ,
S3IT

Consulting & Training

Active Data Management

S3IT

Beteiligung der UZH

Publication & Preservation

HBZ, S3IT

Digital Humanities Lab, Prof. Lukas Rosenthaler, University of Basel

- Setup of a national DLCM platform for Digital Humanities
- Assessment of SALSAAH as possible solution
- Development and testing of needed features
- Support
- Business model



Data Life Cycle Management national: was fehlt?

- **Beratung vor Ort** zu Datenmanagement-Plänen, Einreichung von Forschungsgesuchen, rechtlichen Fragen, bestehenden Möglichkeiten des DLCM, nationalen Diensten etc. (ETH- und EPFL-Bibliothek bieten ihren Forschenden bereits Support)
- Für die **grosse Mehrheit der Forschenden und Projekte (long-tail)**, die nicht durch (inter)nationale Fachinitiativen und fachliche Daten-Repositoryn abgedeckt sind:
Niederschwellige lokale Lösung zum Speichern, Sichern, Teilen, Publizieren und Archivieren ihrer Daten und Datenbanken
- SUK P-2 DLCM und andere nationale Projekte planen keine zentrale fachübergreifende Lösung in diesem Sinne
- Es wird dadurch impliziert, dass Lösungen lokal erstellt werden nach entstehenden (inter)nationalen Standards und best practices und mit gegenseitigem Erfahrungsaustausch. Kleinere Institutionen sollen Dienste bei grösseren kaufen können.
- Beispiel: im Bereich der Langzeitarchivierung (Track 3 von SUK P-2 DLCM) ist noch offen, ob eine zentrale Lösung oder eine Föderation lokaler Plattformen entsteht (ETH hat bereits eigenes Daten-Archiv)





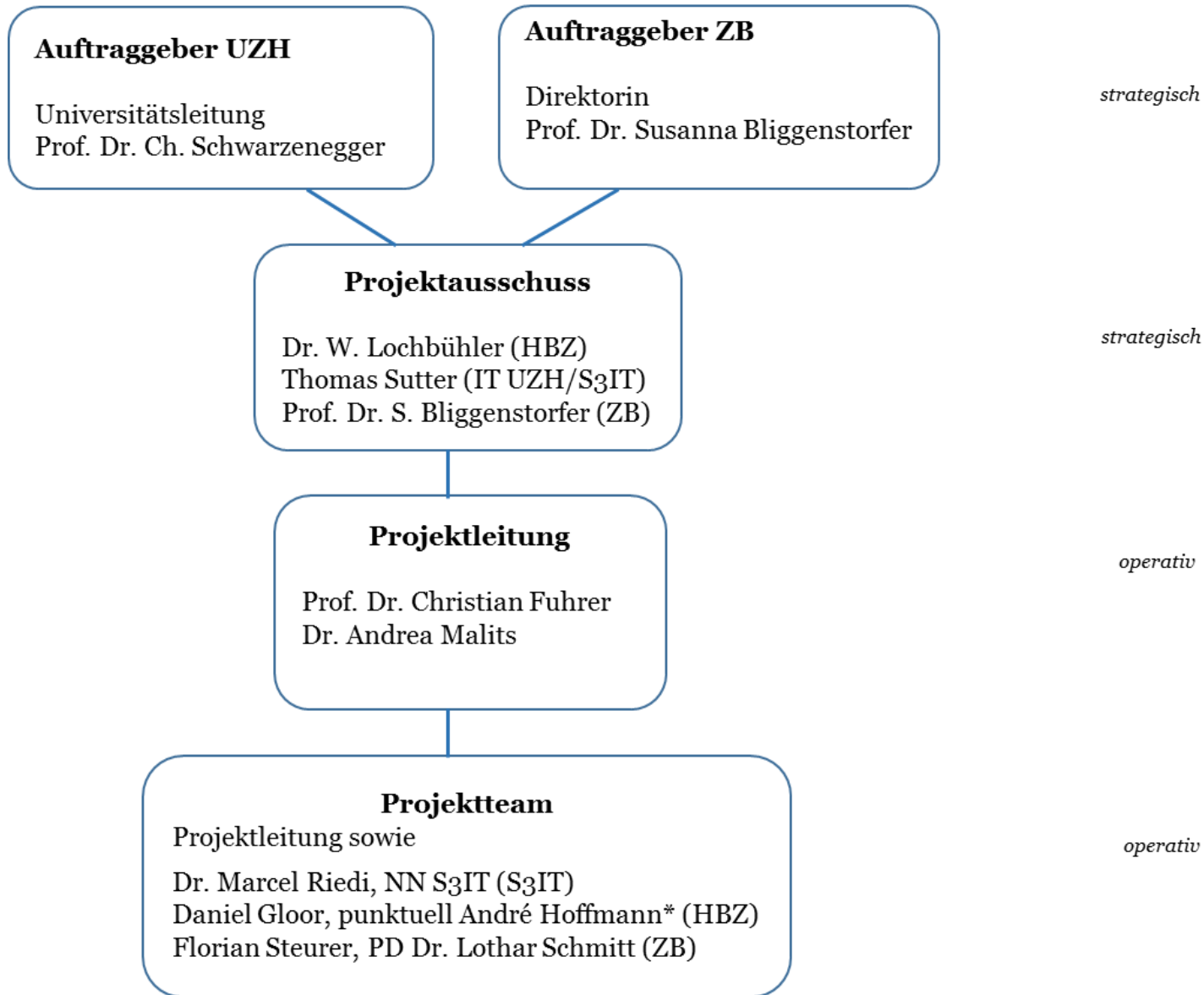
Data-Pilotprojekt an der UZH

- Arbeitsgruppe Forschungsdaten bildete sich 2014, bestehend aus Vertretern der Zentralbibliothek Zürich, S3IT und der Hauptbibliothek der UZH.
- Präsentation eines Konzeptpapiers dieser AG an Direktionen der Hauptbibliothek, Zentralbibliothek, Zentralen Informatik und an Prorektor RWW, Nov. 2014
- Vorprojekt der AG zur Klärung der Bedürfnisse von Forschenden der UZH, Jan. – Sept. 2015
- UL-Antrag für Data-Pilotprojekt bewilligt am 28.4.2016, Laufzeit 1.7.2016 – 31.12.2017
- Kostenrahmen Fr. 350'000
- Weiterführung, Nachhaltigkeit offen

<http://www.news.uzh.ch/de/articles/2015/klarschiff-machen-im-datenwust.html>

http://www.oai.uzh.ch/images/PDFs/Literatur/uzh_journal_2015-3_Archivieren.pdf

Projektorganisation





Ergebnisse des Vorprojekts

Im Vorprojekt erkannte Bedürfnisse von Forschenden der UZH anhand von 17 repräsentativen Gesprächen zum Umgang mit digitalen Forschungsdaten:

- Aufbau eines zentralen Archivs für Forschungsdaten an der UZH inkl. Langzeitarchivierung und abgestufte Zugänge (intern, geschützt, teilbar, öffentlich)
- Migration, Hosting und Pflege bestehender, oftmals selbst entwickelter Datenbanken
- Support und Beratung, auch in verschiedenen rechtlichen Fragestellungen
- Bearbeitung von Anforderungen von Forschungsförderungseinrichtungen (Datenmanagementpläne, Offenheit der Daten/Open Data, Langzeitarchivierung)



Data-Pilotprojekt: Inhalt, Auflagen, Ziel

- Anhand von 4 Forschungsgruppen konkrete, praktische Ansätze zur Umsetzung von FDM prüfen
 - Prof. Dr. Milo Puhan (Institut für Epidemiologie, Biostatistik und Prävention; Patientenstudien)
 - Prof. Dr. Jan Seibert (Geographisches Institut; Hydrologische Daten)
 - Prof. Dr. Christa Dürscheid (Deutsches Seminar, germanistische Linguistik; Variantengrammatik)
 - Digitale Editionen: Prof. B. Naumann, Prof. S. Schneider (Deutsches Seminar); Prof. P. Widmer (Institut für Vergleichende Sprachwissenschaft/Indogermanistik); Prof. T. Weddigen (Kunsthistorisches Institut).
- Kontakte und Abklärungen mit weiteren Forschergruppen nach Möglichkeit
 - Beispiele: Fakultäts-IT der WWF, International Normal Aging and Plasticity Imaging Center (INAPIC) der UZH, ev. weitere
- Testinstallationen und erste Anwendungen von Softwares (Data Repository, Datenplattformen, Tools, ...), Überlegungen zu Schnittstellen
- Berücksichtigung von (nationalen und internationalen) zentralen Lösungsansätzen
- Lokales DLCM in enger Kooperation mit nationalen Lösungen (End Storage). Daher in Ergänzung und Abstimmung mit dem SUK P-2 DLCM Projekt
- Business Case für eine Umsetzung von Forschungsdatenmanagement an der UZH, UL-Antrag



Erste Ergebnisse

Anforderungen der Forschungsgruppen der UZH

- Bestätigung der bekannten Wünsche: Daten speichern, sichern, teilen, publizieren, archivieren
- Eingabe von Daten z.B. durch Doktoranden früh im Forschungsprozess, nicht erst am Ende
- Mehrwert des Nicht-Verlierens von Daten bei Personalwechsel
- Editionen: virtuelle Forschungsumgebung mit Bearbeitungs- und Visualisierungsmöglichkeit, Langzeitarchivierung, nachhaltige Präsentation als Open Data bei abgeschlossenen Editionen, Workflows klären
- Editionen: Zusammenarbeit mit der nationalen Infrastruktur NIE / INE, lokaler Hub und lokale Betreuung gefragt

Andere Institutionen

- Die meisten LERU-Universitäten sowie ETH und EPFL haben bereits eine zentrale Anlaufstelle für DLCM

Plattform-Tests angelaufen

- Evaluation mehrerer Open Source Plattformen
- [Dataverse](#) installiert und konfiguriert: dedizierte Datenplattform und –Repository, Open Source, aktive Community weltweit, national eingesetzt in Holland (DANS), entwickelt an Harvard University



Vielen Dank an...

- **Andrea Malits (Zentralbibliothek Zürich): Zusammenarbeit seit 2014 – AG Forschungsdaten, Data-Pilotprojekt...**
- **André Hoffmann und Martin Brändle (Hauptbibliothek der UZH): SUK P-2 DLCM...**
- **Marcel Riedi (S3IT, ZI UZH); Florian Steurer, Lothar Schmitt (Zentralbibliothek Zürich); Daniel Gloor (Hauptbibliothek der UZH): Data-Pilotprojekt...**
- **...für die tolle Zusammenarbeit**
- **Darum geht es auch bei Forschungsdatenmanagement: verschiedene Stakeholders und Organisationen zusammen gehen das komplexe Thema gemeinsam an und nehmen verschiedene Funktionen wahr**
- **Nachzulesen in der [LERU Roadmap for Research Data](#)**
- **Sie für Ihre Aufmerksamkeit**
- **FRAGEN?**