



Forschungsdaten: Von den Grenzen der Langzeitarchivierung und dem Wert des Datenmanagements

6. Kongress Bibliothek & Information Deutschland Leipzig, 14. März 2016

Dr. Matthias Töwe, ETH-Bibliothek, ETH Zürich





Gliederung

- OAIS und Forschungsdaten
- Datenmanagement und Langzeitarchivierung
- Kompetenzen
- Verantwortlichkeiten
- Lernprozess
- Fazit





OAIS und wir

Bei uns...

- OAIS breit akzeptiert in Gedächtnisinstitutionen
- Die Referenz zum Aufbau vertrauenswürdiger Langzeitarchive
- Als Bezugsrahmen bewährt

Data Management

...und sonst?

Access

Unbekannt bei Forschenden und Partnern z.B. aus der IT





OAIS und Forschungsdaten

Zur Erinnerung:

- OAIS macht keine Annahme zur Art der Daten…
- ...wurde ausgehend von Forschungsdaten konzipiert...
- ...durch eine Arbeitsgruppe von Weltraumagenturen...
- ...zur Adressierung erkannter Schwachstellen...
- ...bei der langfristigen Erhaltung ihrer Daten

Access

Archival Storage





OAIS und Forschungsdaten

Charakteristik der Daten aus Weltraumforschung und Raumfahrt:

- Einmalige Daten mit immensem Erzeugungsaufwand
- Spezifischer Hintergrund mit wenigen Akteuren
- Enge Beziehung von Datenproduzenten und Agenturen
- Domänen von Erfassung, Sharing und zentralem Datenarchiv weniger stark getrennt als z.B. in der Hochschulforschung

Archival Storage





Domänenmodell «Weltraumforschung»

Private Research, Shared Research, Publication, and the Boundary Transitions Migration Process Migration Process **Shared Domain Public Domain Private Domain** Authorised by research team leader Performed by research team IT Collaboration support/IR staff Institutional Support System document System/Research (Plone, TWiki, repository Sharepoint) Private Domäne entfällt May link to data objects stored in May link to data objects stored in weitgehend Involves object selection, migration, assignment of persistent identifier, creation of fixity metadata, augmentation of Collaboration other metadata Public Data Data Store Store/Repository Publication Curation Boundary This domain involves the core research This domain involves researchers This domain involves the public sphere team as they undertake the research, outside the core team as they (publication in the sense of making usually within a single institution. collaborate with colleagues, often public). Access usually open to all. Access is often tightly controlled as across institutions. Access is more hypotheses and analyses are developed. Version 1.4.3, http://andrew.treloar.net/, 19 Mar 2012 open, but not everything is shared.





OAIS und «Small Data» / «Long Tail»?

- Andere Ausgangslage...
- ...aber die grundlegenden Konzepte gelten genau so!

Zusätzliche Herausforderungen für OAIS-konforme Lösungen:

- Heterogenität der Daten, Methoden, Standards und Erwartungen
- Kein gemeinsamer Nenner wenig übertragbare Annahmen
- Akteure des Lebenszyklus' von Daten agieren oft noch getrennt
- «Ingest» zeitlich und organisatorisch entfernt von Produktion





Domänenmodell «Small Data» / «Long Tail»

Private Research, Shared Research, Publication, and the Boundary Transitions Migration Process Migration Process **Shared Domain Public Domain Private Domain** Authorised by research team leader Laboratory Performed by research team IT Collaboration Information Management support Support System (Plone, TWiki, System/Research Management Sharepoint) System Öffentliche Domäne ist May link to data objects stored in weiterhin die Ausnahme May link to data objects stored in Involves object selection. migration, assignment of access controls, augmentation of metadata Research Data Collaboration Store Data Store Collaboration Curation Boundary This domain involves the core research This domain involves researchers This domain involves the public sphere team as they undertake the research, outside the core team as they (publication in the sense of making usually within a single institution. collaborate with colleagues, often public). Access usually open to all. Access is often tightly controlled as across institutions. Access is more hypotheses and analyses are developed. Version 1.4.3, http://andrew.treloar.net/, 19 Mar 2012 open, but not everything is shared.





Beispiel «Designated Community»

Im Modell:

- Baustein des Konzepts
- Hilfskonstrukt zur Eingrenzung von Ziel, Art und Umfang
- Wirkt sich auf die Definition der Information Packages aus

Small Data / Long Tail:

- Viele spezifische Communities mit unbekannten Bedürfnissen...
- ...schon gar nicht als Projektion in die Zukunft





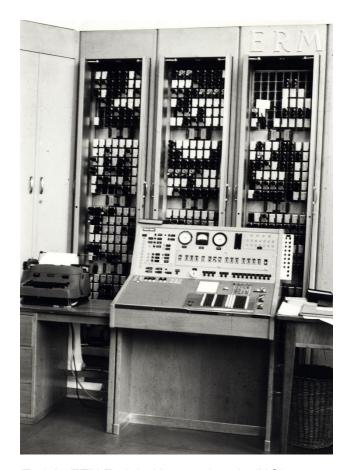
Beispiel «Information Package»

Im Modell:

- Selbst dokumentierendes Informationspaket
- Abhängigkeiten nur zu Elementen, die ihrerseits archiviert sind

Small Data / Long Tail:

- Abhängigkeiten von Hardware schwer abzugrenzen:
 Forschende heben Hardware auf und hoffen
- Teilweise subtile Abhängigkeiten z.B. von Dateiattributen
- Auf welche Communities hin muss Paketinhalt ausgelegt sein?



Zürich, ETH Zürich, Hauptgebäude (HG), Institut für angewandte Mathematik, ERMETH-Rechenautomat ETH-Bibliothek Zürich, Bildarchiv / Fotograf: Unbekannt / Ans_00290 / CC BY-SA 4.0





Trennung der Verantwortlichkeiten...

«Forschende machen Forschung»

- Was gehört dazu?
- Für viele ist Datenmanagement integraler
 Bestandteil des Forschungsprozesses im Sinne guter wissenschaftlicher Praxis
- Ob, was und wie es getan wird, ist aber extrem unterschiedlich
- Bewusste Entscheidung für sehr einfache Lösungen kann adäquat sein



Laboranten in Raum F108, 2.7.1954 ETH-Bibliothek Zürich, Bildarchiv / Fotograf: Unbekannt / Ans_00069 / CC BY-SA 4.0





...führt zu Missverständnissen

«Erst machen wir Forschung und dann archiviert Ihr nach OAIS»

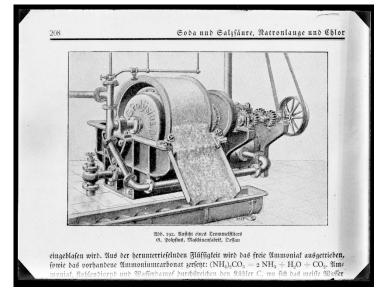
- Entspricht in vielen Fällen den organisatorischen Strukturen (Forschungsbereiche vs. Infrastruktureinrichtungen)
- Verkennt Notwendigkeit, durch Datenmanagement Daten überhaupt archivierbar zu machen
- Erzeugt die falsche Vorstellung, Verantwortung für den eigenen Beitrag zu Datenmanagement und –erhaltung sei delegierbar





Datenmanagement als Filter

- Wo reicht vorübergehende Aufbewahrung?
- Wo ist Aufwand zur Vorbereitung der langfristigen Erhaltung gerechtfertigt?
- Wo sind Schwierigkeiten absehbar, wenn Daten viel später an ein Langzeitarchiv übergeben werden sollen?
- Wessen Kompetenzen sind gefordert, um sie zu beheben oder zumindest die Auswirkungen abzumildern?
- Wer ist verantwortlich für welche Massnahmen?
- NB: Heutige Entscheidungen können sich als falsch erweisen!
 - Alles «irgendwie» aufzubewahren, ist keine Option



Ansicht eines Trommelfilters, G. Polysius, Maschinenfabrik Dessau ETH-Bibliothek Zürich, Bildarchiv / Fotograf: Photographisches Institut der ETH Zürich / PI_28-C-0090 / Public Domain Mark



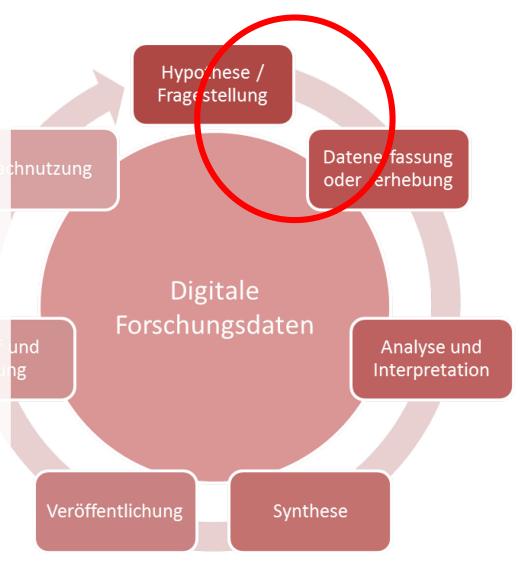
Ziel

 Forschende bereits bei der Planung des Datenmanagements unterstützen und kritische Entscheidungsfähigkeit verbessern

 Daten als Investitionsgut sehen, nicht als Verbrauchsmaterial

 Adressaten: Doktorandinnen, Postdocs, technisch Verantwortliche, IT-Support

Professorinnen: Eher zur Wahrnehmung ihrer
 Verantwortung ermutigen > Hochschulpolicy, gute
 wissenschaftliche Praxis

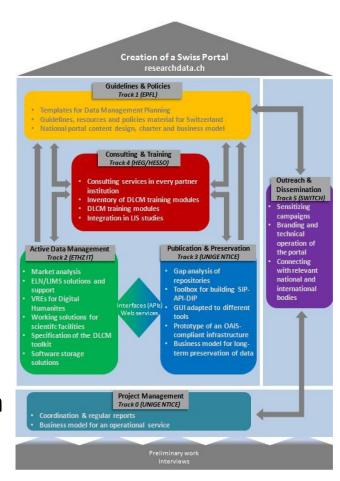






Und wie?

- «One size fits all?» Nein, sicher nicht!
- «Prêt-à-porter!» Baukasten aus Angeboten von Schulungen bis Tools, um breitere Kreise zu erreichen
- Kooperationen nutzen!
- In der Schweiz vor allem im Projekt «Research Data Life-Cycle Management» (DLCM) von swissuniversities
- Austausch von Vorhandenem und abgestimmte Erarbeitung von neuen Materialien, Tools, Services



DLCM, E. Blumer, P.-Y. Burgi





Lernprozess – oder Quadratur des Kreises?

- Verstehen der konkreten Wünsche von Forschenden und Abgleich mit den Anforderungen der Langzeiterhaltung...
 - «Hier habt Ihr die Daten macht doch mal, was Ihr für richtig haltet»
 - «Fasst die Daten bloss nicht an ich brauche die genau so zurück» (Keine Veränderung von Dateiattributen, kein Ersetzen von Sonderzeichen in Dateinamen o.ä.)
- ...und den Aufwand vertretbar und angemessen halten
- OAIS bleibt auch dabei ein valider und wichtiger Bezugsrahmen!





Fazit

- Datenmanagement und Langzeitarchivierung sind zwei Seiten einer Medaille...
- ...zwischen denen immer noch zu wenig Durchlässigkeit besteht
- Aber es bewegt sich an vielen Orten etwas! Siehe Präsentationen der nächsten Tage...

Offen:

- Wie viel Datenmanagement ist notwendig, um die «richtigen» Daten langfristig für eine Nachnutzung bewahren zu können?
- Wie viel Ressourcen darf Langzeitarchivierung binden, um auch eine noch ungewisse Nachnutzung zu ermöglichen...
- ...ohne neue Forschungsmöglichkeiten zu hemmen?





Herzlichen Dank!

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

<u>matthias.toewe@library.ethz.ch</u> <u>http://www.library.ethz.ch/Digitaler-Datenerhalt</u>

