ETH zürich

ETH Library

Workflows für Forschungsdaten

Aktuelle Anwendungsfälle

Conference Poster

Author(s):
Gastl, Angela

Publication date:
2015

Permanent link:
https://doi.org/10.3929/ethz-a-010380326

Rights / license:
In Copyright - Non-Commercial Use Permitted

This page was generated automatically upon download from the ETH Zurich Research Collection.
For more information, please consult the Terms of use.

Workflows für Forschungsdaten: Aktuelle Anwendungsfälle

Wieviel Lang-zeitarchivierung darf es sein?

Angela Gastl, Fachstelle Digitaler Datenerhalt
ETH Zürich, ETH-Bibliothek, 8092 Zürich, Schweiz
angela.gastl@library.ethz.ch, www.library.ethz.ch

Ausgangslage

Auftrag der ETH-Schulleitung, ein Archivierungsangebot für Forschungsdaten aufzubauen: **ETH Data Archive**

«Teilarchiv»

- ETH-Technologietransferstelle möchte ein Teilarchiv innerhalb des ETH Data Archive aufbauen
- Ziel ist die Archivierung von Software Code (entstanden an der ETH Zürich)
- Datenpaket enthält neben dem Code auch Kontextinformation
- Nicht alle Daten sind frei zugänglich

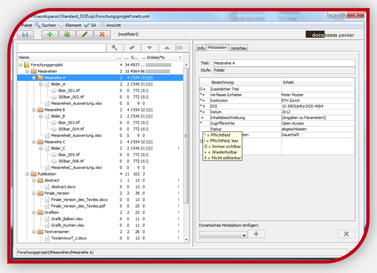
«Emeritus»

- Emeritierender Professor hat div. Datenträger mit Daten zu Dissertationen («Diss-Archive»)
- Diss-Archive sollen für 10 Jahre gesichert werden
- Zugriff nur für Emeritus
- Dissertationstext ist bereits auf ETH-Dokumentenserver zugänglich (inkl. DOI)

«Academic Supervisor»

- Forschungsprojektleiter/Academic Supervisor ist gemäss ETH-Richtlinie verantwortlich für Datenmanagement und -Archivierung über das Projektende hinaus
- Projektleiter setzt sich ein für einheitliche Datenhaltung im Hinblick auf Archivierung

Datenaufbereitung



Pre-Ingest-Tool «docuteam packer»

«Teilarchiv»

- Datenproduzent bereitet Datenpaket gem. Vorgaben der ETH-Transferstelle auf (1 ZIP-Container mit unkomprimierten Files)
- Datenproduzent erhält Log-in für ETH Data Archive
- Datenproduzent lädt Datenpakete manuell hoch und erfasst deskriptive Metadaten

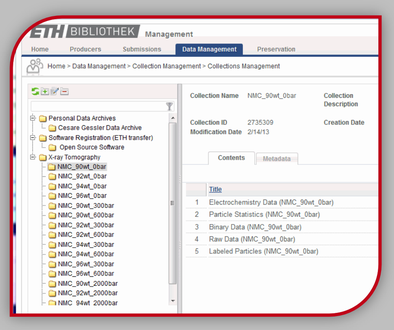
«Emeritus»

- Emeritus packt Daten pro Datenträger vorgängig in ZIP-Container (unkomprimiert)
- Emeritus erhält Log-in für ETH Data Archive
- Emeritus lädt Datenpakete manuell hoch und erfasst deskriptive Metadaten

«Academic Supervisor»

- Forschungsgruppe verwendet ein Pre-Ingest-Tool
- Pre-Ingest-Tool unterstützt Daten-strukturierung, -beschreibung und -vorvalidierung (inkl. DOI-Vergabe)
- Archivwürdige Daten werden per Knopfdruck ins ETH Data Archive hochgeladen

ETH Data Archive



Das ETH Data Archive basiert auf dem Digital Preservation System «Rosetta»

«Teilarchiv»

- Daten werden beim Ingest wenn möglich validiert und in der originalen Form archiviert
- Daten durchlaufen inhaltliche Prüfung (Approval) durch ETH-Transferstelle
- Daten werden Collection zugeordnet
- Nach 15 Jahren erfolgt Review durch Fachstelle und Hochschularchiv (Option: Löschung)

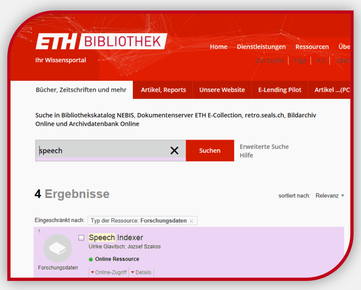
«Emeritus»

- Daten werden beim Ingest entpackt, validiert und wieder in einen ZIP-Container gepackt
- Nicht-validierbare Formate werden trotzdem in Originalform archiviert
- Daten werden der Emeritus-Collection zugeordnet («Personal Data Archives»)
- Nach Ablauf einer 10-jährigen Frist werden die Daten gelöscht («Expiry Date»)

«Academic Supervisor»

- Automatischer Ingest ab Pre-Ingest-Tool
- Ordnerstruktur bleibt als Collection- und Sub-Collectionstruktur erhalten
- Datenformate werden soweit möglich validiert
- Daten werden unbefristet aufbewahrt

Zugriff



Recherche im Wissensportal der ETH-Bibliothek («Primo»)

«Teilarchiv»

- Metadaten aller Datenpakete sind über das Wissensportal der ETH-Bibliothek auffindbar (ETH Data Archive als Sammlung in Primo)
- Frei zugängliche Daten erhalten DOI
- Download des Datenpakets als ZIP
- Falls Inhalt geschützt ist, Log-in erforderlich

«Emeritus»

- Emeritus erhält einen persistenten Link auf seine Kollektion mit allen Archivpaketen
- Download des Datenpakets als ZIP
- Zugriff über Internet mit Login (View, Download)
- Zugriff ist beschränkt auf Emeritus

«Academic Supervisor»

- Metadaten der Archivpakete sind über das Wissensportal der ETH-Bibliothek auffindbar
- Persistenter DOI-Link direkt auf Daten
- Daten sind über Rosetta-Viewer zugänglich
- Falls Inhalt geschützt ist, Log-in erforderlich

FAZIT

- ➔ Langfristig heisst nicht immer ewig archivieren
- ➔ Formatvalidierung bei Forschungsdaten sehr schwierig
- ➔ Von Individuallösungen für Pilotpartner hin zu Standard-Dienst-leistungen für viele
- ➔ Qualitativ gute deskriptive Metadaten sind schwierig zu erreichen

«Teilarchiv»

- Ein festgesetzter Review der Daten nach 15 Jahren (mit Option Löschung) erlaubt es, den späteren Aufwand zur Datenpflege zu reduzieren
- Datenproduzenten agieren weitgehend selbständig
- Aufwand ETH-Transferstelle beschränkt sich auf die Freigabe (Approval) der neu hochgeladenen Datenpakete

«Emeritus»

- Das «digitale Daten-Schliessfach» ermöglicht dem emeritierten Professor, auch von ausserhalb der Hochschule seiner Datenaufbewahrungspflicht korrekt (und elegant) nachzukommen
- Ein einfacher, standardisierter Workflow soll den Service für diese Kundengruppe möglichst niederschwellig gestalten
- Die befristete Datenaufbewahrung entspricht dem Kundenbedürfnis und verringert den Pflegeaufwand (Digital Curation)

«Academic Supervisor»

- Das Pre-Ingest-Tool unterstützt das Datenmanagement einer Forschungs-gruppe, indem archivierungswürdige Daten im laufenden Prozess selektiert und aufbereitet werden können
- Das ETH Data Archive gewährleistet den langfristigen und – wenn gewünscht – freien Zugang zu den Daten und fördert so die Nachnutzung