Conference Poster

Workflows for Research Data
Current Use Cases

Author(s):
Gastl, Angela; Töwe, Matthias

Publication Date:
2015

Permanent Link:
https://doi.org/10.3929/ethz-a-010362436

Rights / License:
In Copyright - Non-Commercial Use Permitted
Workflows for Research Data: Current Use Cases

**Starting point**
- Mandate from ETH Board to implement an archiving service for Research Data: ETH Data Archive

**Data processing**
- Pre-Ingest-Tool «docuteam packer»

**ETH Data Archive**
- ETH Data Archive is based on the Digital Preservation System «Rosetta» (Ex Libris)

**Access**
- Search in ETH Zurich’s Knowledge Portal («Primo»)

**Conclusion**
- Long term is not necessarily forever
- Format validation is very difficult for research data
- From individual solutions to pilot partners to standard services for many
- Good quality descriptive metadata are hard to obtain

- **Emeritus**
  - Retiring professor has offline media with data from doctoral theses («Diss-Archives»)
  - Diss-Archives are to be safeguarded for 10 years
  - Access for the professor only
  - Text of doctoral theses is already online on ETH E-Collection (incl. DOI)

- **Sub-Archive**
  - ETH Technology Transfer Office want to set up a sub-archive within ETH Data Archive
  - The aim is to archive software code originating from within ETH Zurich and licensed to third parties
  - Data package contains code and context information
  - Not all data is freely accessible

- **Emeritus**
  - Emeritus packages data per medium (e.g. CD or DVD) into ZIP-container (uncompressed)
  - Emeritus receives log-in for ETH Data Archive
  - Emeritus uploads data package manually and enters descriptive metadata

- **Academic Supervisor**
  - According to ETH guidelines, project leader / academic supervisor is responsible for data management beyond end of project
  - Project leader advocates uniform data management with respect to archiving

- **Sub-Archive**
  - Data producer prepares data package according to guideline of ETH Technology Transfer Office
  - Data producer receives log-in for ETH Data Archive
  - Data producer uploads data package manually and enters descriptive metadata

- **Academic Supervisor**
  - Research group uses the Pre-Ingest-Tool
  - Pre-Ingest-Tool supports data structuring, description and pre-validation (incl. DOI-creation)
  - Selected data to be archived are submitted to ETH Data Archive by one click

- **Sub-Archive**
  - Data is validated upon ingest where possible and archived in its original format
  - Data go through a check (approval) by ETH Technology Transfer Office
  - Data is assigned to collections
  - After 15 years, a review is envisaged by ETH Technology Transfer Office and University Archives of ETH with the option of deletion

- **Emeritus**
  - Upon ingest, data is unpacked, validated and re-packed into a ZIP-container
  - Formats which cannot be validated are nevertheless archived in their original format
  - Data is assigned to the Emeritus’s collection («Personal Data Archives»)
  - Data will be deleted after a retention period of 10 years has expired («Expiry Date»)

- **Emeritus**
  - Emeritus receives a persistent link to his collection with all data packages
  - Data package can be downloaded as ZIP
  - Internet access with login (view, download)
  - Access is restricted to Emeritus

- **Academic Supervisor**
  - Automatic ingest out of the Pre-Ingest-Tool
  - Folder structure is preserved as a structure of collections and sub-collections
  - File formats are validated as far as possible
  - Data is archived with an unlimited retention period

- **Sub-Archive**
  - Metadata of all data packages can be found in ETH Zurich’s Knowledge Portal (ETH Data Archive acts as one source for Primo)
  - Open Access data are assigned a DOI
  - Data package can be downloaded as ZIP
  - If access is restricted, a log-in is required

- **Emeritus**
  - Emeritus receives a persistent link to his collection with all data packages
  - Data package can be downloaded as ZIP
  - Internet access with login (view, download)
  - Access is restricted to Emeritus

- **Academic Supervisor**
  - Metadata of archival packages can be found in ETH Zurich’s Knowledge Portal («Academic Supervisor»)
  - Persistent DOI-link points directly to data
  - Data are delivered via viewers from Rosetta
  - If access is restricted, a log-in is required

- **Sub-Archive**
  - A defined Review of data after 15 years (with option of deletion) permits to reduce data maintenance in the future
  - Data producers act rather autonomously
  - Effort for ETH Technology Transfer Office is reduced to approval of newly submitted data packages

- **Emeritus**
  - The «digital safe deposit box» enables a retiring professor to fulfill his obligations for data safeguarding even from outside ETH correctly (and elegantly)
  - A straightforward, standardized workflow should lower barriers to use the service for this group of customers
  - The limited safeguarding of data caters for a concrete customer need and reduces the need for data maintenance (Digital Curation)

- **Academic Supervisor**
  - The Pre-Ingest-Tool supports data management within a research group by facilitating the ongoing selection and processing of data to be archived as part of daily business
  - ETH Data Archive guarantees the long term and – if possible - open access to data and thus encourages their re-use