

Long-term preservation: Integrating DSpace and Rosetta at ETH Zurich

Conference Poster

Author(s):

Hirschmann, Barbara ; la Roi, Andreas ; Scowen, Greg

Publication date:

2019-06-11

Permanent link:

<https://doi.org/10.3929/ethz-b-000344235>

Rights / license:

[Creative Commons Attribution 4.0 International](#)

Long-term preservation: Integrating DSpace and Rosetta at ETH Zurich

Barbara Hirschmann (orcid.org/0000-0003-0289-0345)¹, Andreas la Roi (orcid.org/0000-0003-4209-1361)¹, Greg Scowen (orcid.org/0000-0002-1824-3399)¹

¹ETH Library, ETH Zurich

1 ETH Zurich & ETH Library

ETH Zurich

- Founded 1855
- One of the world's leading universities for science and technology
- 26'700 students and staff from over 120 different countries
- 11th in THE ranking, 7th in QS ranking
- 12'000 scholarly publications per year

Publication & preservation services

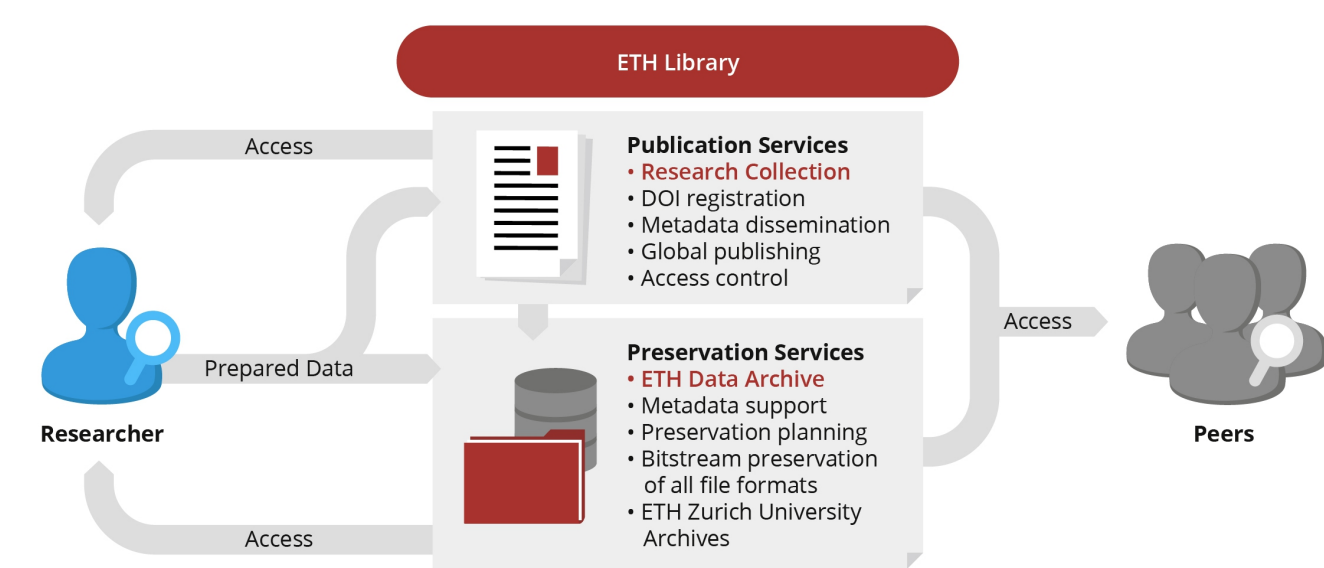


Fig. 1. Schematic overview of publication and preservation services provided by ETH Library

2 DSpace & Rosetta

Research Collection (DSpace)

- A "3 in 1" repository
 - open access repository
 - institutional bibliography
 - research data repository

Data Archive (Rosetta)

- Used as a «dark archive» for Research Collection content

Preservation workflow

- All items that reached the final quality status "QS6" are pushed as SIPs (Submission Information Packages) to Rosetta
- Export includes files and metadata
- Format identification and validation upon ingest (excluding content of ZIP containers)
- Continuous AIP (Archival Information Package) updates for manual changes to Research Collection items
- Yearly AIP updates for «automatic updates» (e.g. metadata changes produced by update scripts)

Example export numbers

	Apr 2019	Mai 2019
New Items	1'202	800
Updates	22'386	19'020

3 Preservation workflow

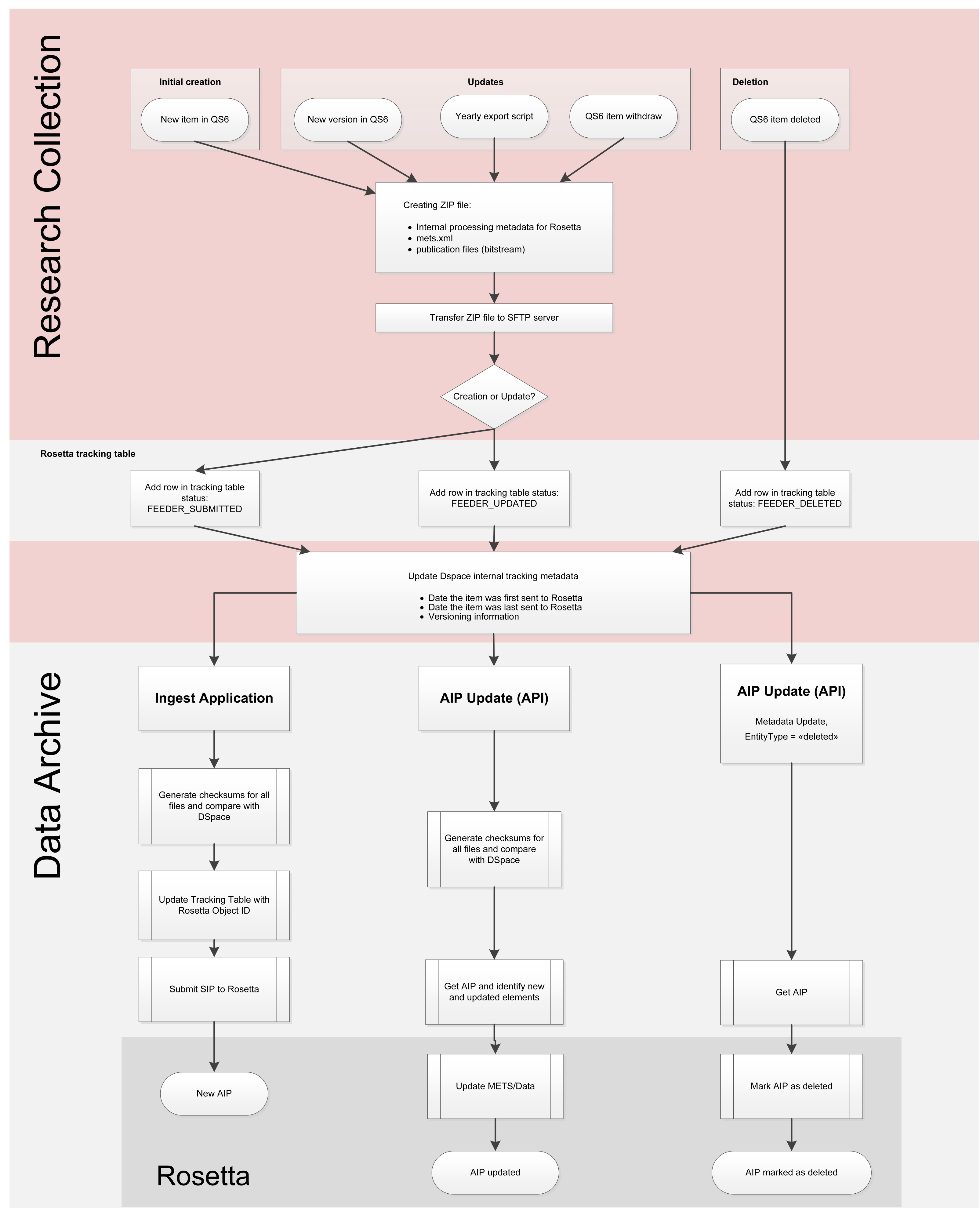


Fig. 2. Preservation workflow between DSpace and Rosetta

4 Archived content in Rosetta

- 196'390 items, 266'978 files, 864 GB of data
- Mostly XML (METS files) and PDF formats, PDFs are validated with JHOVE
- Up to now more than 430'000 updates, resulting in multiple item versions in Rosetta
- Preservation planning will be done for limited number of file formats only

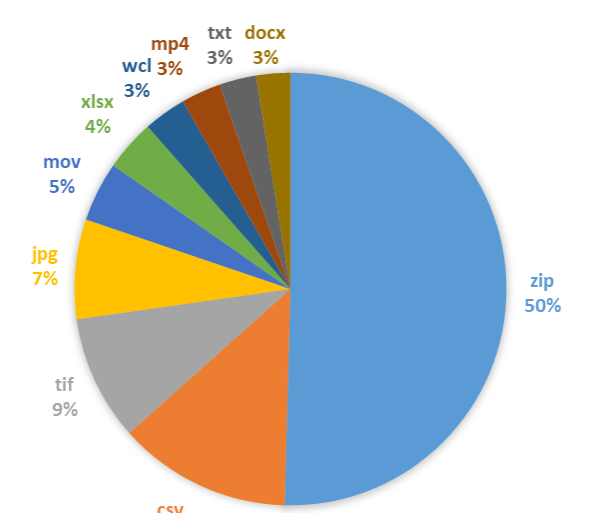


Fig. 3. Formats other than XML and PDF (top ten)

