Out of sight but not out of mind: Archivematica in the cloud

Tim Hutchinson – ACA Regina, 13 June 2015
BACKGROUND
Previous digital preservation efforts

- Transfer from legacy media where possible – but not systematic
- Very selective normalization
- Some documentation especially for larger scale transfers
- Hard to be proactive about accepting born digital content
MEMORANDUM

To: Dean Peter Nikiforuk, College of Engineering
From: Robert N. Kavanagh, Associate Vice-President
Date: Wednesday, December 17, 2014
Subject: Applications to Apple Canada Education Foundation, Engineering and Computing Science Division

There seems to be a need to clear up a few points about submissions to the ACEF. I hope that the following can be of use to you and your faculty.

1) The Apple Canada Education Foundation exits to assist development of educational applications of Apple products. This implies that the Foundation does not support projects that are exclusively of a research nature; but research projects that have a significant educational component are received favourably. It also implies (by the word ‘assist’) that the support from the ACEF is likely to be only part of the costs of a project. In very rough terms, the ACEF looks for roughly matching ‘hard dollar’ contributions from the University or from other sources that a faculty member might have developed.

2) The ACEF is not in the business of funding teaching labs. It might however contribute to the development of pilot educational projects for labs which are subsequently justified within and supported by the institution.
Organizational context

- University Archives separate from Library (sort of) until 2007
  a) 2007: Arms-length unit reporting to library dean
  b) 2010: “normal” library unit
  c) 2013: Merged with Special Collections
  d) 2014: Digital Projects added to UA&SC umbrella

- IT support provided by Library all along

- But support for digital preservation more likely if it benefits the Library too
The hook: digital projects

- Archives and Library – approx. 50 digital projects since 1995 (mostly led by Archives)
  a) Approx. 2 TB

- Saskatchewan History Online – three-year provincial project led by library (2012-2014)
  a) By end of project, 4TB+
  b) Plus 10TB (and counting) - Courtney Milne site

- Consultation with Artefactual, early 2012

- Need same infrastructure to preserve digitized content or born digital
ARCHIVEMATICA AS A SERVICE: THE DECISION
Rationale

- Testing Archivematica since version 0.8 – production installation (as of 1.0) already planned
- Major digitization project – library IT resources stretched
- Similar storage costs as for local hosting
- Easy exit strategy
Rationale cont’d

- Support from Artefactual
- Extra features e.g. fixity checking
- Contribution to COPPUL capacity
- Also considered support contract (local install)
ARCHIVEMATICA AS A SERVICE: THE STORY SO FAR
Testing and deployment

- Training and sandbox launch: Apr. 2014
- Extensive testing
- But then a slow route to production. Detours included:
  a) Islandora/Archivematica integration
  b) pre-Archivematica workflow
  c) Artefactual schedule; newly uncovered bugs
- Production launch: Dec. 2014
- Kept separate pipeline for ongoing testing
Islandora integration
1. Content Upload
   1.1) (Drupal cron + Fedora content validation) used to trigger content upload to Archivematica
   1.2) One upload per Fedora Object
   1.3) Metadata includes an AIP ID (generated from user input or based on collection)
   1.4) Islandora saves EM-IRI from 201 response

2. Create Transfer
   2.1) Parse & validate METS.xml
   2.2) Identify files to request
   2.3) Create Transfer object
   2.4) Pass file URI’s to asynchronous process

3. Collect Files
   3.1) Retrieve files from Fedora REST API
   3.2) Confirm checksums

4. Transfer & Ingest
   4.1) Move files to watchedDirectory
   4.2) Complete Transfer & Ingest (manual or auto)
   4.3) Delete Hi-Res datastream from Islandora (if selected as config option)

5. Finalize Package (AIP)
   Islandora needs to know when the 'last' object in AIP has been uploaded, determined automatically or based on input

6. Check Archivematica Status
   6.1) Request Status of Transfer

7. Status Response
   7.1 Archivematica returns info on current state of Transfer [at which stage(s)??]

8. Content deleted?
   note: content update not possible until Archivematica versioning enhancements in release 1.1 (Spring 2014)

9. Log Islandora deletion
   9.1) Mark objects as deleted from access system
   9.2) Update search index
   NOTE: Does not trigger Archivematica AIP delete.
Pre-Archivematica workflow

- Some transfers easier to put straight into Archivematica
- Legacy media
  - Incrementally building digital forensics workstation
- Challenge of appraisal
End of year one

- U of S has renewed for another year of the COPPUL Archivematica service
- Very good technical support from Artefactual - particularly to minimize downtime
- Additional group training
- Cost reduced to reflect limited use of “tier 3” support tickets
Future considerations

- Difficult to shift completely away from network drive storage – current limitations in DIP structure
  - a) Not yet using AtoM for born-digital records
- Dominance of Microsoft Office files – normalization not automated
- Local cost of storage/backups may change – affecting cost effectiveness of hosted service
CONCLUSIONS
Archivematica in the “cloud”

- Storage/security – inherent trust of another university
  a) Further investigation of security policies may be needed before including university records
- Canadian servers
- More accessible support than typical cloud services
- Management of records not really outsourced – just backend
- But also doesn’t fit “pay as you go” model
“It might not be perfect, but we have a plan...”

U of S College of Commerce (Hangar Building), 1951.
University Archives photo A-3057