Digital Preservation at the University of Arizona Libraries

By Christine Kollen

University of Arizona Libraries

September 5, 2017

Outline

- Introduction
- National Digital Stewardship Alliance - Levels of Preservation
- Digital preservation at the UA
- Arizona’s LiveData project
- Rosetta – digital preservation infrastructure
- Going forward

Introduction

- University of Arizona Libraries is developing digital preservation strategy
- As in many libraries, digital preservation is considered important but resources and lack of knowledge limit what we can do
- Key departments in the library involved with digital preservation
  - Office of Digital Innovation and Stewardship (ODIS)
  - Technology Strategy & Services (TeSS)
  - Special Collections

Digital preservation

"…series of managed activities necessary to ensure continued access to digital materials for as long as necessary."[1]

"…digital preservation is about crafting the right approach for a given preservation context. … Digital preservation requires the work of craftspeople who can reflexively approach digital preservation problems in situ and develop approaches that match the resource, material, and conceptual constraints of a given setting."[2]

Illustration by Jørgen Stamp digitalbevaring.dk CC BY 2.5 Denmark

1. Digital Preservation Handbook, 2nd Edition, http://handbook.dpconline.org/, Digital Preservation Coalition © 2015.
2. Owens, Trevor. Theory and Craft of Digital Preservation, Draft of 4th Chapter. (page 4-1) *http://www.trevorowens.org/2017/06/full-draft-of-theory-craft-of-digital-preservation/*

National Digital Stewardship Alliance (NDSA) - Levels of Preservation

	Level 1 (Protect your data)	Level 2 (Know your data)	Level 3 (Monitor your data)	Level 4 (Repair your data)
Storage & Geographic Location	•2 complete copies, not collocated •Move data on heterogeneous media to storage system	•At least 3 complete copies •At least 1 in a different location •Document storage system/media	•At least 1 copy in different location •Obsolescence monitoring process for storage system/media	•At least 3 copies in locations w/different disaster threats •Comprehensive plan
File Fixity & Data Integrity	•Check file fixity on ingest (if provided) •Create fixity info if not provided	•Check fixity on all ingests •Use write-blockers w/original media •Virus-check high risk content	•Check fixity at fixed intervals •Maintain logs; supply audit on demand •Virus-check all content	•Check fixity of all content in response to certain activities/events •Ability to replace/repair corrupted data •Ensure no one has write access to all copies

NDSA Levels of Preservation (continued)

	Level 1 (Protect your data)	Level 2 (Know your data)	Level 3 (Monitor your data)	Level 4 (Repair your data)
Information Security	•Identify who has authorization to individual files •Restrict who has those authorizations	Document access restrictions for content	Maintain logs of who performed what actions	Perform audit of logs
Metadata	•Inventory of content & storage location •Ensure back-up and non-collocation of inventory	•Store administrative metadata •Store transformative metadata and log events	Store standard technical/descriptive metadata	Store standard preservation metadata
File Formats	Encourage use of limited set of open formats & codecs	Inventory of file formats in use	Monitor file format obsolescence issues	Perform format migrations, emulations etc. as needed

Status of the UA's digital preservation

NDSA Levels of Preservation – in general, we fall mainly in Level 1

The ISO 16363 standard (Trusted Data Repositories) checklist has several areas that are applicable to digital preservation, can be used to assess where you are at

- Do we have the needed staff? → Recently hired staff with digital preservation, data management, and development expertise
- Do we have sustainable funding → Not at this time
- Do we have a digital preservation policy → Not currently, our new staff will help to develop
- Do we have processes that ensure our digital objects are reliable and authentic over time → Not currently, our new staff will help to develop

Digital preservation

Components of digital preservation

- Systems, e.g. Archivematica, Preservica, Rosetta etc.
- Storage, e.g. Arkivum, LOCKSS
- Schemas, e.g. PREMIS, METS
- Digital forensics stations and workflows – on pause, new staff will reassess and restart

Research Environments

- Data preservation environments like iRODS data grids or DataVerse
  - The UA CyVerse uses iRODS, has a number of services including some data preservation
- Consulting with digital preservation services like DuraCloud, Chronopolis etc. – for high risk or high value digital files

Live-Data: Managing Data for a 21st Century Research Enterprise

Arizona Tri-University project

- Arizona State University
- University of Arizona
- Northern Arizona University

Multi-year project with multiple components, funded by Arizona Board of Regents' Innovation Fund

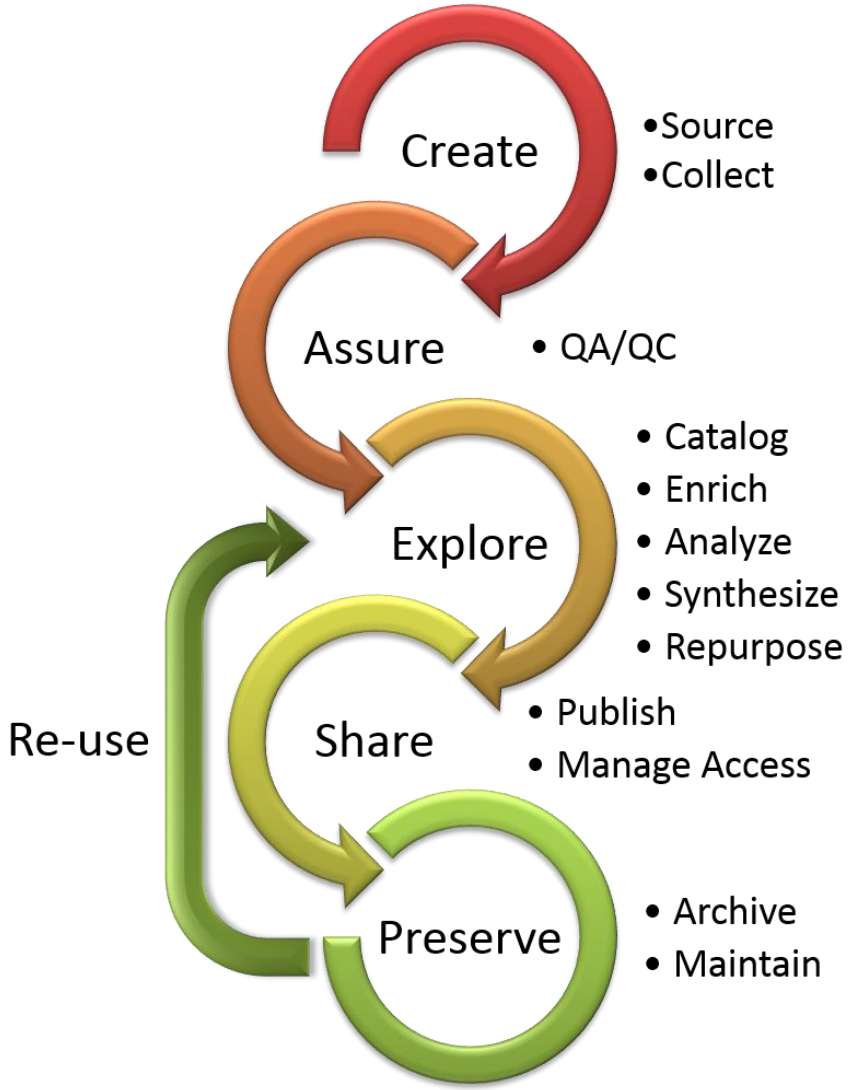
- COS Pivot - database for identifying funding opportunities
- Research Profiles – implemented Elsevier's SciVal Experts, integrated with Pure to become Pure Experts. No longer has the same functionality, we are exploring alternative products
- ResearchAZ – focused on access and digital preservation

ResearchAZ

- SHARE – collaborating with Center for Open Science to implement cross-repository searching
  - Two ASU repositories were harvested as part of a pilot

- Ex Libris Rosetta – digital preservation infrastructure to support long-term curation of research outputs (both data and scholarship)

Research Data Life Cycle



LiveData project addresses the Share → Preserve → Reuse portions of the research data lifecycle

SHARE promotes research outputs from across the three state universities

Rosetta – enables efficient long-term Tri-University research data and scholarly outputs

Ex Libris Rosetta

- Capable of supporting ingestion of large numbers of digital objects
  - Batch loads
  - Staging Areas
  - Multiple deposit hierarchies
- Flexibility to configure to meet archiving and preservation needs of all three universities, including workflows to support individual institutional needs
- Enable all three universities to promote long-term discovery, sharing and re-use of scholarly digital collections, including research data

UA Rosetta Implementation Group Goals

- Create workflows in Rosetta – ingest and management of digital collections
- Develop plan to migrate legacy digital collections into Rosetta
- Develop policies, processes and workflows to address digital preservation in Rosetta

- In addition, actively participate in tri-university implementation – identify opportunities for shared policies, processes and governance of Rosetta

Rosetta Issues

- Does not have a robust North American community
- Not open source, you can’t look “under the hood”
- Requires custom scripting for each content type ingested into the system
- Not interoperable with key systems (e.g. ArchiveSpace) and infrastructure (e.g. AWS)
- Does not have an intuitive interface through the ingest and processing workflows

Going forward – components in place

LiveData project funding is ending

- The LiveData project helped the Libraries:
  - develop closer working relationships with ASU and NAU
  - provided us with additional experience and a chance to experiment
  - demonstrated the need to hire additional staff for digital preservation both on ODIS and TeSS
- UA, ASU, and NAU will continue to collaborate
- Project charter with Tri-University members is being developed on roles, changes and near-term action plan

Going forward (continued)

- Permanent ODIS Department Head and Metadata Librarian started last year
- Digital Preservation Librarian started in July
- Research Data Management Specialist starts in September
- Two new technologist approved and hired (on TeSS)
  - Digital preservation DevOps staff member
  - Digital preservation Developer staff member
- Looking at another product – Archivematica
    - Open source
    - Active user community

Conclusion

UA Libraries is beginning to develop digital preservation services

- Rosetta project provided us with the opportunity to experiment
- We have established good working relationships with ASU and NAU and have plans to continue to collaborate
- Demonstrated need for additional staff
  - ODIS – librarians and specialists
  - TeSS – development professionals
- We are excited to be moving forward!

Resources

"Digital Preservation Handbook". 2nd Edition. *Digital Preservation Coalition*. July 10, 2017. http://handbook.dpconline.org

National Digital Stewardship Alliance. "NDSA Levels of Preservation". *Library of Congress.* July 10, 2017. http://www.digitalpreservation.gov:8081/ndsa/activities/levels.html

Owens, Trevor. "Theory and Craft of Digital Preservation", *Trevor Owens. June 30, 2017.* *http://www.trevorowens.org/2017/06/full-draft-of-theory-craft-of-digital-preservation/*

Primary Trustworthy Digital Repository Authorisation Body Ltd. "Preparing for an Audit". *International Organization for Standardization*. July 10, 2017. http://www.iso16363.org/iso-certification/preparation/

Questions?

Contact

- Chris Kollen, kollen@email.arizona.edu or 520-305-0495