



ACADEMIC
PRESERVATION TRUST

Learning to Run

University of Cincinnati Libraries path to
contributing content to APTrust

Digital Collection Environment

- 3 digital repositories
 - DSpace -- UC Digital Resource Commons
 - LUNA -- UC Libraries Image and Media Collections
 - Hydra -- Scholar@UC Institutional Repository
- 1 website
- 1 Storage Area Network file system



Workflow Considerations

- Preservation vs. Recovery
 - platform-specific bags or platform-agnostic bags?
 - derivative files
- Bagging level
 - record level or collection level?

UC Libraries Workflow Choices

- Recovery is as important as preservation
 - Include platform specific files and derivatives
- Track platform provenance
 - use bag naming convention or additional fields in bag-info.text
- Track collection membership
 - use bag naming convention or additional fields in bag-info.text

Bagging (crawling)

1st iteration

- Start with DSpace, Cincinnati Subway and Street Improvements collection
- Replication Task Suite to produce Bags
 - one bag per record
- Bash scripts to transform to APTrust Bags
- Bash scripts to upload to test repository
- Worked well, but...

Bagging (walking)

2nd iteration

- Same starting place
- Export collection as Simple Archive Format
 - ~ 8,800 records, ~ 375 GB
- Create bags using bagit-python
 - 250 GB is bag-size upper limit

Bagging (walking)

2nd iteration (continued)

- Collection had three series
 - two series under 250 GB
 - each series was put into a single bag
 - one series over 250 GB
 - used a multi-part bag to send to APTrust, APTrust merges split bags into one
- Bash script to upload bags to production repository

Bagging (running)

3rd iteration

- Testing of tools created by Jamie Little of University of Miami
- Jamie can tell you more but...

Long-Term Strategy (marathons)

- Bag for preservation and recovery
- For most content
 - one bag per collection
 - except for Scholar@UC
 - manual curation and bagging
 - except for Scholar@UC
- Use bag-info.txt to capture platform provenance and collection membership

Why are we doing this again?

- Libraries' and Archives' historic commitment to long-term preservation is the *raison d'être*, a core justification, for our support of repositories. It's why our stakeholders trust us and will place their content in our repositories.
- We expect APTTrust to be our long-term preservation repository to meet our responsibilities to our stakeholders.

Why are we doing this again?

- We know from some experience, and reports from our peers, that backup procedures, while still essential, are fraught with holes.
- We expect APTTrust to be our failover if any near-term restore from backup proves incomplete.

UC Libraries, Digital Collections and Repositories

Linda Newman
Head, Digital Collections
and Repositories
linda.newman@uc.edu

Nathan Tallman
Digital Content Strategist
nathan.tallman@uc.edu