A Micro-Services-Based Approach for Curation and Preservation Solutions

Stephen Abrams
Patricia Cruse
John Kunze
Perry Willett

University of California Curation Center
California Digital Library
Who are we?
What do we do?

- Innovative curation solutions (preservation and access) for the UC community and external partners
  - EZID, Merritt repository, Web Archiving Service (WAS)
  - Guidelines and best practices
  - Integration with other CDL programs (e.g., OA publishing, special collections) and external initiatives (e.g., DataCite, DataONE, HathiTrust, etc.)
That’s simple, right?

- Ever increasing number, size, and diversity of content
That’s simple, right?

- Ever increasing number, size, and diversity of content
- Expanded use by new constituencies
That’s simple, right?

• Ever increasing number, size, and diversity of content
• Expanded use by new constituencies
• Ongoing (potentially disruptive) changes in technology and user expectation
• Increasing obligations, shrinking budgets

Merritt repository

“I know my desktop content is at risk; what should I do?”

“How can I ensure persistent availability?”

“What’s a good way to share the data underlying a recent publication?”

“How can I meet the data management requirements of my grant?”
Merritt repository

Dark archive for preservation masters

Bright archive for preservation and end-user access

Preservation back-end for existing discovery services

Integration with distributed data grids
Managerially friendly
Model free
Strongly versioned

Agnew, Duncan C., Intensities for the Four Largest Shocks of the New Madrid Earthquake Sequence

- **primary identifier**: ark:/13030/qt2sc7s4wb
- **title**: Intensities for the Four Largest Shocks of the New Madrid Earthquake Sequence
- **creator**: Agnew, Duncan C.
- **date**: 2011-04-29
- **local id**: 
- **last modified**: 2011-05-02 09:57 am
- **created**: 2011-04-29 05:32 pm
- **total size (of versions)**: 84.8 KB
- **total versions**: 2

Download object

Current Version

Version 2: 2011-05-02 09:57 am
- mrt-erc.txt
- qt2sc7s4wb.bp_meta.xml
- qt2sc7s4wb.meta.xml
- qt2sc7s4wb.pdf

Prior Versions

Version 1: 2011-04-29 05:32 pm
Easy submission

Add Object

Merritt > Collection: Demo Merritt collection > Add Object

Upload objects directly from your local drive or submit a manifest to add objects posted on a web server.

Object Lookup

Search object, author, title, date or identifier.

Add Object Help

Mouse over the icon for details about each element in the form.

The following guides provide detailed instructions for adding objects and related tasks such as preparing manifest files.

- Adding Objects
  - A manifest
  - Basic metadata
  - Objects
  - More data

Optional description:

You will be able to search by the following fields:

- Title:
- Creator:
- Date:
- Local ID:
Merritt micro-services

- Merritt is built from a micro-services toolkit
  - IdM/Authn/Authz 💩 LDAP ➔ GhOST/Shibboleth
  - Persistent identifiers 💩 EZID
  - Persistent storage 💩 CAN/Pairtree/Dflat/Checkm/ReDD
  - Fixity 💩 Fixity ➔ Version 2
  - Replication 💩 Replication
  - Catalog 💩 Inventory/4store
  - Ingest 🎈 Ingest/Zookeeper
  - Characterization 🐘 JHOVE2
  - Discovery 🐘 XTF
  - Transformation 🐘
  - Notification 🐘
  - Annotation 🐘
Micro-service choreography
Micro-services

The Unix philosophy

- “Make each program do one thing well”
- “To do a new job, build afresh rather than complicate old programs by adding new features”
- “Expect the output of every program to become the input to another, as yet unknown, program”
- “Design and build software … to be tried early”
- “Don't hesitate to throw away the clumsy parts and rebuild them”

L. McIlroy et al.
“Unix time-sharing system forward”

- Complex emergent behavior
- Low barrier, low maintenance, low commitment
- Policy neutral, protocol/platform independent
- The file system is the database

http://www.flickr.com/photos/oskay/265899811
Questions?

For more information

UC Curation Center
http://www.cdlib.org/uc3
uc3@ucop.edu

Merritt repository
http://merritt.cdlib.org/

UC3
Stephen Abrams  Margaret Low
Lisa Colvin  David Loy
Patricia Cruse  Mark Reyes
Scott Fisher  Tracy Seneca
Alex Genadinik  Joan Starr
Erik Hetzner  Marisa Strong
Greg Janée  Perry Willett
John Kunze