Agenda

• Who we are
• Data curation, publication, and sharing
• Tools to help you
  – DMPTool
  – DataUp
  – Dash
  – WAS
• Summary
• Discussion
Who we are

“To support the University of California community’s pursuit of scholarship and ... public service mission”
Data curation, publication, and sharing

• Increasingly, a requirement for funding and publication
• Transparency ↔ trust
• Reduce needless duplication of effort
• Leverage prior investments
• Expand the reach of your research, and get credit for it
• Good for science, good for scientists
Data curation, publication, and sharing

• Create/acquire a dataset in a form that is inherently preservable and (re)usable
• Describe the dataset in scientifically-meaningful ways
• Give the dataset a unique identifier for persistent citation
• License the dataset under CC0 or CC-BY
• Deposit the dataset in a (non-commercial) repository where it will receive pro-active curation management
• Expose the dataset for harvesting by abstracting/indexing services and search engines
DMPTool

“Fulfill institutional and funder mandates”

dmptool.org
blog.dmptool.org
github.com/CDLUC3/dmptool/wiki
DMPTool

• Free and open for all
• Hosted by CDL, with code released as open source
• Supports data management requirements for NSF, NIH, NEH, NOAA, IMLS, and other federal agencies and private funders
• New version released on May 29
• Developed by a partnership of universities, museums, and researchers, with support from Sloan Foundation and IMLS

dmptool.org
blog.dmptool.org
github.com/CDLUC3/dmptool/wiki
• In addition to fulfilling external requirements, the DMPTool provides:
  – Framework to plan for management of research data
  – Comprehensive list of issues involved with data management best practices
  – Information about local resources and services: repositories, workshops, consultation services, etc.
  – Community of stakeholders: researchers, lab managers, IT specialists, archivists, grant administrators, funding agencies

dmptool.org
blog.dmptool.org
github.com/CDLUC3/dmptool/wiki
DataUp

“Curation for tabular datasets”

dataup.org
dataup.cdlib.org
Excel is often the database of choice for research.
DataUp

- Drag-and-drop data upload
- Opportunity to add descriptive metadata
- Assignment of persistent identifier / generation of persistent citation
- Best practices check
- Packaging and submission to ONEShare repository

Performed automatically

dataup.org
dataup.cdlib.org
“Long-term identifiers made easy”
“Long-term identifiers made easy”

- store citation metadata for identifiers in a variety of formats
“Long-term identifiers made easy”

- EZID (easy-eye-dee) makes it easy to create & manage:
  - create identifiers for anything: texts, data, bones, terms, etc.
  - store citation metadata for identifiers in a variety of formats
  - update current URL locations so citation links are never broken
  - use EZID's programming interface for automated operation at scale
  - choose from a variety of persistent identifiers, including ARKs and DataCite DOIs
  - try out EZID anytime and contact us for more information

No more 404 errors!
• “Long-term identifiers made easy”

DOI for persistent citation and bi-directional linking between publications and underlying data
• “Long-term identifiers made easy”
“Preservation and access”

- No prescriptive requirements on content genre, type, format, structure, or metadata
- Strong versioning maintains complete change history
- Restricted or public access – under your control
- Enforceable data use agreements (DUAs)
- Storage replication to UCLA and UCSD, with ongoing auditing
- Integration with EZID and DataONE
- Proactive preservation analysis, planning, and intervention

Merritt is a new cost-effective repository service from the University of California Curation Center (UC3). Use Merritt to preserve, manage, share, and access your valuable digital content. Merritt is built upon a Micro-Services approach to digital curation based on developing independent, but interoperable, services that embody curation values and strategies. Services are small and self-contained, and collectively easier to develop, deploy, maintain, and audit. This design approach is detailed in the UC3 Curation Foundations paper.

Learn more about Merritt:
- Handout: Merritt Uses and FAQ
- Merritt User Guide

Need a Merritt Account?
Fill out the Contact UC3 form or email uc3@ucop.edu.

Merritt is a service of the University of California Curation Center. Copyright © 2000-2014 The Regents of the University of California. Privacy Policy | Image Credits
DataONE

• “Data observation network for Earth”

- Cyberinfrastructure
  - Distributed grid of member and coordinating nodes
  - Aggregated discovery
  - Investigator’s toolkit

- Community

[Image of DataONE search tool and map]

dataone.org
“Data sharing made easy”
• Preservation repositories are complex systems
• Far too often, their interfaces are complicated and meant only for IT professionals and archivists
• Dash provides a set of user-friendly screens to step through the process:
  – Select/upload files associated with a dataset
  – Augment with descriptive metadata
  – Review that the dataset meets requirements and is ready
  – Submit to the Merritt preservation repository with optionally replication to DataONE
Dash

• Upload dataset files
• Add descriptive information
• Review the dataset
• Submit to a repository

Record contains all recommended fields
- Go Back
- Edit Description

Uploaded Files
- Record must include at least one file
- Upload
- More Files

Ready to go?
- Submit to Dash
Dash

- Search/browse and discovery
• “Capture and preserve the web”
The web is a volatile environment
• WAS captures and preserves important web content
• WAS captures the web over time
WAS provides curators with tools to capture the free web:

- Schedule web crawls on regular or customized basis
- Focus on website itself or include linked sites
- Brief 1-hour or full 36-hour crawls
- Analyze results with a range of reports
- Search across captured websites
- Keep archive restricted, or provide public access
- Fee-based service
WAS

• WAS includes archives based on events:
  – 2003 California recall election
  – 2007 Southern California wildfires

• Thematic archives:
  – Grateful Dead archives
  – US Labor unions and organizations
  – California political blogs

• Comprehensive archives of web-domains:
  – Emory University
  – University of Michigan

was.cdlib.org
webarchives.cdlib.org
Service takeaways

- **DMPTool**  
  Create data management plans required by funders or journals using campus resources

- **DataUp**  
  Curation services tailored for tabular datasets

- **Dash**  
  Simplified interfaces for repository submission and discovery

- **EZID / Merritt**  
  Core infrastructural services generally hidden beneath simple intuitive interfaces

- **WAS**  
  Curation services tailored for web-published content and data
Summary

- Good data management practice is critical to the success of the academic enterprise and scholarly advancement
- Management solutions should be integrated into existing research systems and workflows
- The UC Libraries are a natural partner for data management advice and solutions
- UC3 offers a comprehensive roster of innovative and intuitive curation services applicable across the data and scholarly lifecycle
For more information

- **UC Curation Center**  [www.cdlib.org/uc3](http://www.cdlib.org/uc3)  [datapub.cdlib.org](http://datapub.cdlib.org)  [uc3@ucop.edu](mailto:uc3@ucop.edu)
- **DMPTool**  [dmptool.org](http://dmptool.org)
- **DataUp/ONEShare**  [dataup.org](http://dataup.org)
- **Dash**  [datashare.ucsf.edu](http://datashare.ucsf.edu)
  - **EZID**  [ezid.cdlib.org](http://ezid.cdlib.org)
  - **Merritt**  [merritt.cdlib.org](http://merritt.cdlib.org)
  - **DataONE**  [dataone.org](http://dataone.org)
- **WAS**  [was.cdlib.org](http://was.cdlib.org)