

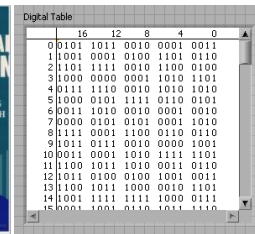
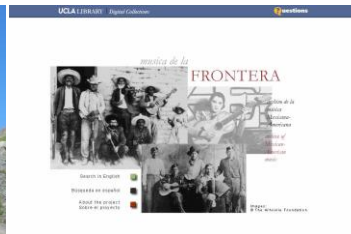


UC Curation Center

Building Services, Building Communities, Supporting Data Intensive Research

January 2012 – International Data Curation Conference

Patricia Cruse, Director
University of California Curation Center
California Digital Library



California Digital Library

Serving the University of California

- 10 campuses
- 360K students, faculty, and staff
- 100's of museums, art galleries, observatories, marine centers, botanical gardens
- 5 medical centers
- 5 law schools
- 3 National Laboratories

CDL supports the research lifecycle

- Collections
- Digital Special Collections
- Discovery & Delivery
- Publishing Group
- **UC Curation Center (UC3)**



California Digital Library (CDL)



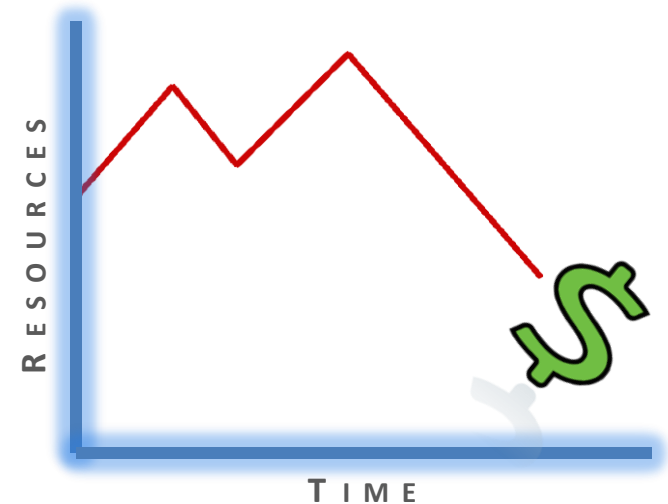
Our environment circa 2002-2008

Focus on preservation
For memory organizations
Content: museum & library
Infrastructure: static
Services: hosted
Sustainability: ?



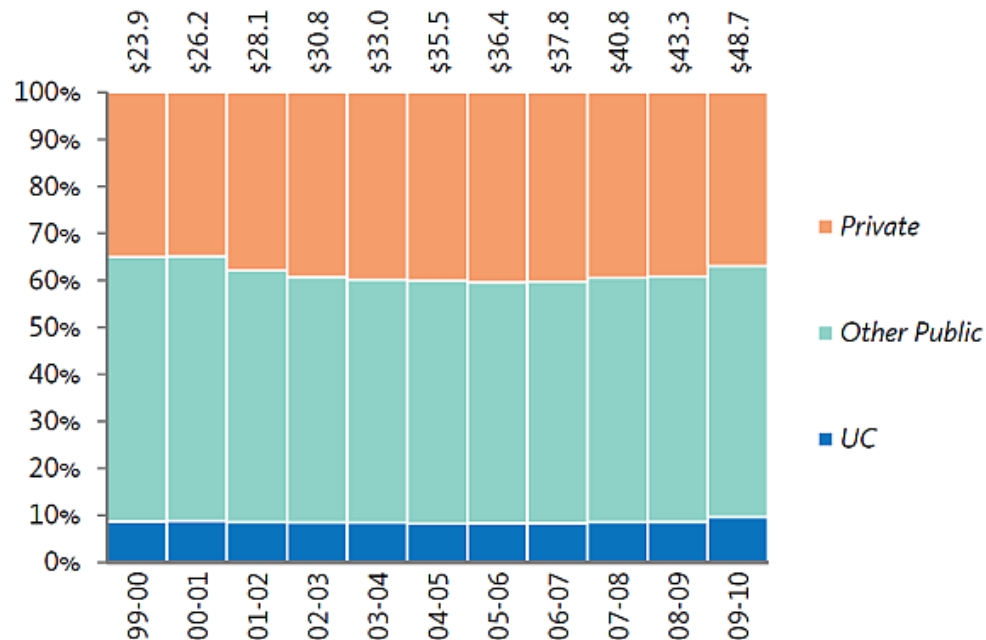
The changing landscape

- Ever increasing number, size, and diversity of content
- Ever increasing diversity of partners, and stakeholders
- Decreasing resources
- Inevitability of disruptive change
 - Technology
 - Institutional mission



The University of California's role in research

UC share of U.S. research expenditures, Universitywide, 1999-2000 to 2009-10



The University of California performs about one-twelfth of all the academic research and development conducted in the United States.

Source: IPEDS

What keeps users up at night?

How much will it cost?

How do I fulfill the data management requirements of my grant?

Why should I care about preservation? I just need a place to put my data.

Where can I get help?

Are there standards or best practices I should be aware of?

How can I share my work with my colleagues?

What are "metadata"?

How can I provide access to my work?

Can't my work be included in the Web of Science?

How can I make sure I get credit?

How can I publish the data associated with my publications?



Our environment since 2008

Focus on preservation

For memory organizations

Infrastructure: static

Services: hosted

Content: museum & library

Sustainability: ?

→ *curation* (lifecycle)

→ and now data producers

→ applications enabling
diverse use

→ partnered, self-serve

→ research, web crawls

→ new cost models

Pivot and work differently: build community and applications

- ✓ Engage with the research community early in the lifecycle
- ✓ Prioritize the use of scarce resources to support emerging initiatives
- ✓ Start small: provide simple solutions that build up to solve more complex problems
- ✓ Deploy simple and flexible infrastructure and services that can be used in diverse ways
- ✓ Develop partnerships and collaborate now more than ever!

Five services to meet diverse needs

1. Create, edit, share, and save data management plans



2. Create, share, archive, publish tabular data



3. Create and manage persistent identifiers



4. Curation repository: store, manage, preserve, and share research data



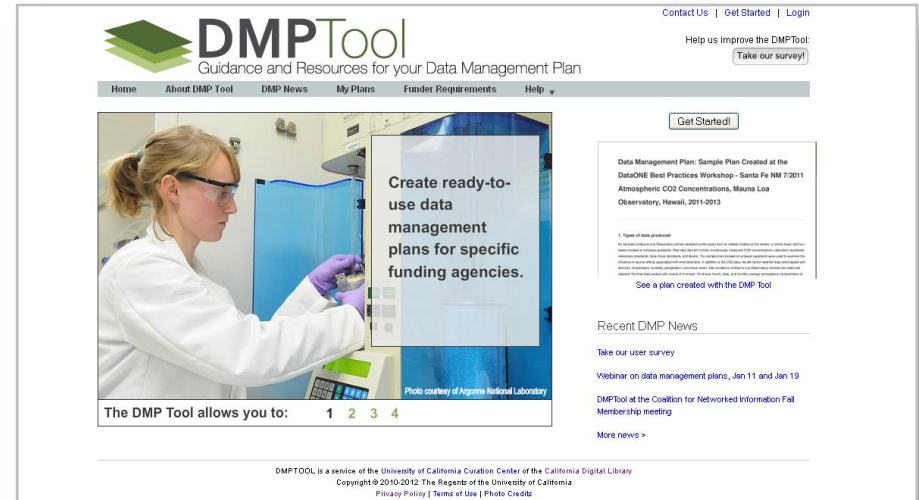
5. A service to collect, manage and preserve Web-published content



Service 1: DMPTool

Meeting funding agencies data management plan requirements

- Connect researchers to resources to create a data management plan
- NSF and directorates, NIH, NEH, IMLS, foundations plus
- Customizable

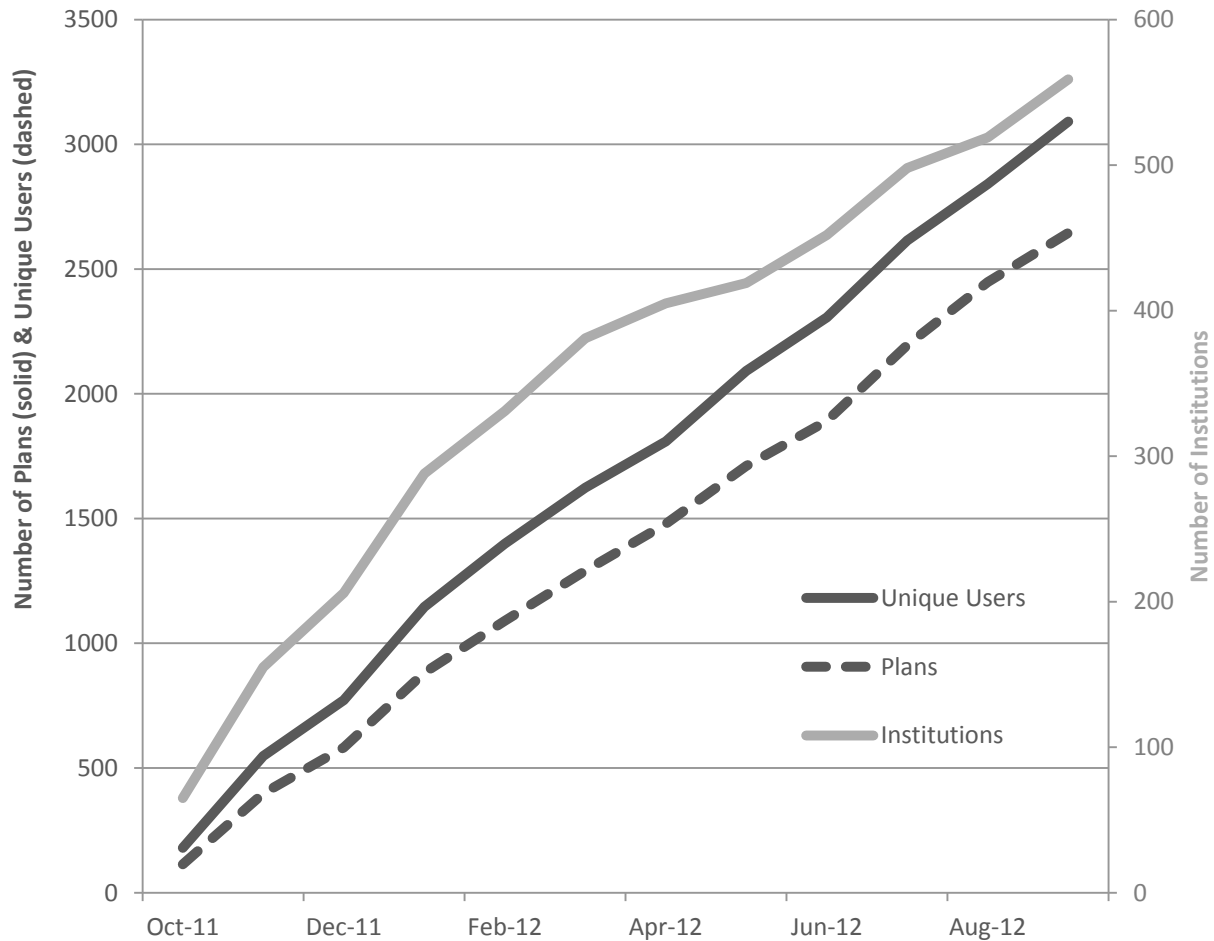


Primary Functions

1. Step-by-step “wizard”
2. Templates and examples
3. Links to institutional resources and agency information
4. Plan publication and sharing



DMTool usage



New DMPTool development

- Funding from Sloan Foundation
- Starts now
- Partners:
 - UC3
 - DataONE
 - University of Virginia Library
 - University of Illinois Urbana Champaign
 - DMPTool Partners
- Enable institutions to easily shape and exploit the DMPTool to meet local need
- Provide support for the full data management life cycle
- Foster the emergence of an engaged open source community of users and developers
- Enable collaboration and connect stakeholders to each other and institutional resources, including services, expertise and guidance.

Goal: DMPTool can provide a centralized point for the consolidation, sharing and dissemination of expertise and advice in good data management practices and solutions

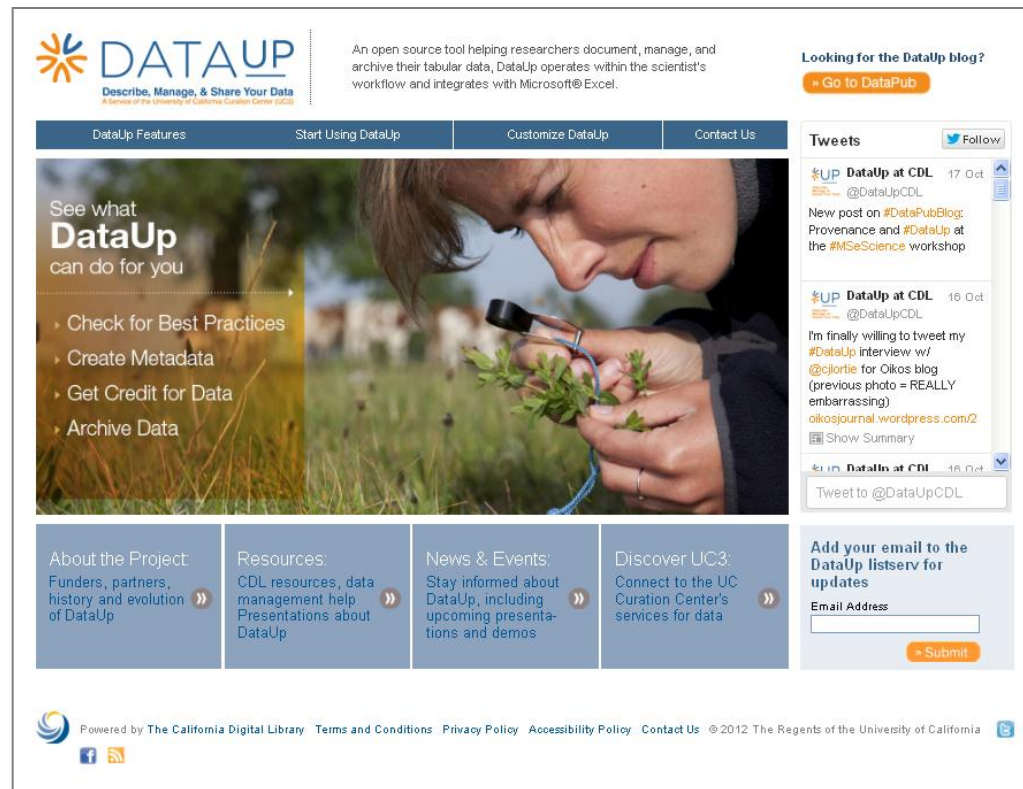
Service 2: DATAUP

Describe, Manage, & Share Your Data
A Service of the University of California Curation Center (UC3)

Create, share, archive, connect, publish data
Excel is the database of choice for many researcher

Primary Functions

1. An Excel 1) add-in and 2) cloud application
2. Document data
3. Check for good data practices
3. Obtain identifier and citation
4. Archive and share



The screenshot shows the DataUp website interface. At the top left is the DataUp logo with the tagline "Describe, Manage, & Share Your Data" and "A Service of the University of California Curation Center (UC3)". To the right, a text box states: "An open source tool helping researchers document, manage, and archive their tabular data, DataUp operates within the scientist's workflow and integrates with Microsoft® Excel." Below this is a navigation bar with links: "DataUp Features", "Start Using DataUp", "Customize DataUp", and "Contact Us". The main content area features a large image of a person using a magnifying glass on a plant, with the text "See what DataUp can do for you" and a list of features: "Check for Best Practices", "Create Metadata", "Get Credit for Data", and "Archive Data". To the right of the main content is a "Tweets" section with two tweets from @DataUpCDL. At the bottom of the main content area are four columns: "About the Project" (Funderson, partners, history and evolution of DataUp), "Resources" (CDL resources, data management help, Presentations about DataUp), "News & Events" (Stay informed about DataUp, including upcoming presentations and demos), and "Discover UC3" (Connect to the UC Curation Center's services for data). Below these columns is an email subscription form with the text "Add your email to the DataUp listserv for updates" and a "Submit" button. At the very bottom of the page, there is a footer with social media icons, the text "Powered by The California Digital Library", and links for "Terms and Conditions", "Privacy Policy", "Accessibility Policy", and "Contact Us". The footer also includes the copyright notice "© 2012 The Regents of the University of California".

Microsoft
Research

GORDON AND BETTY
MOORE
FOUNDATION

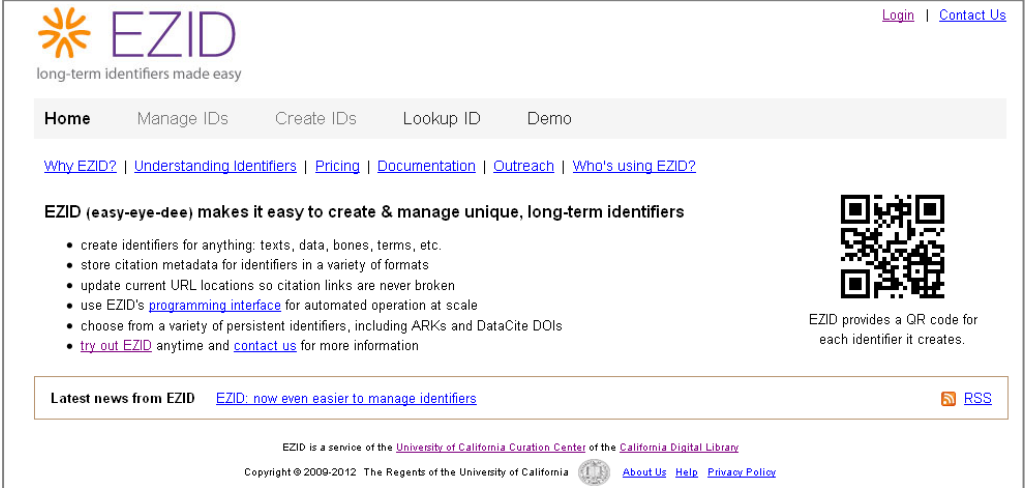
 UC3

University of California
 CDL
California Digital Library

Service 3: EZID

Long term identifiers made easy

- Credit to data producers and data publishers
- A link from the traditional literature to the data
- Cost model in place to support infrastructure
- Range of customers



The screenshot shows the EZID website homepage. At the top left is the EZID logo (an orange starburst) and the text "EZID long-term identifiers made easy". In the top right corner, there are links for "Login" and "Contact Us". Below the logo is a navigation menu with "Home", "Manage IDs", "Create IDs", "Lookup ID", and "Demo". Underneath the menu is a list of links: "Why EZID?", "Understanding Identifiers", "Pricing", "Documentation", "Outreach", and "Who's using EZID?". The main content area features the heading "EZID (easy-eye-dee) makes it easy to create & manage unique, long-term identifiers" followed by a bulleted list of features: creating identifiers for various formats, storing citation metadata, updating URL locations, using a programming interface, and choosing from various identifier types like ARKs and DataCite DOIs. To the right of this list is a QR code with the text "EZID provides a QR code for each identifier it creates." Below the main content is a "Latest news from EZID" section with a link to "EZID: now even easier to manage identifiers" and an RSS icon. At the bottom, it states "EZID is a service of the University of California Curation Center of the California Digital Library" and includes a copyright notice for 2009-2012 along with links for "About Us", "Help", and "Privacy Policy".

Primary Functions

1. Create identifiers (DOIs and ARKs) for anything
2. Citation metadata and persistent citation links
3. DOIs from DataCite
4. Global discovery via Primo and (soon) Web of Knowledge



Service 4: Merritt



- Curation repository open to the UC community and beyond
- Discipline / content agnostic
- Architecture allows for diversity

Primary Functions

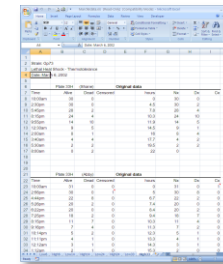
1. Deposit
2. Manage (metadata, versions, etc)
3. Access (expose)
4. Share (with other researchers)
5. Preserve

Merritt's diverse service offering: meeting a range of needs

- Dark archive for important digital assets
-



- Bright archive with direct discovery and access
-



- Preservation back-end for existing or new discovery and content management systems
-



- Integration with distributed data grids
-



Merritt example: UC Berkeley's Research Hub back-end preservation



University of California Berkeley

Welcome to the Research Hub

Create

10 GB free (1000s of files)
Collaborate with colleagues everywhere



Organize and enrich

Add folders, comments, tags and descriptions
Locate files quickly with thumbnails, previews and advanced search
Communicate with blogs, wikis, forums and social networking



Keep safe

Control who can view, edit or comment
Access from anywhere – even from your smartphone



[Learn more](#)

Sign In

Cal Net ID:

Passphrase:

Use your CalNet ID to access Research Hub

Research Hub News

[Research Hub Feature Highlight #1](#)

Did you know that you can drag and drop multiple files and folders from your computer straight into the Research Hub? This is possible by setting up a feature called [WebDAV](#). Here's how: [Windows, Mac, Cyberduck](#).

[NEW! Research Hub gets even better](#)

We have upgraded the Research Hub. New features include:

- Drag & Drop file upload
- Audio/Video playback
- Photoshop & Illustrator previews
- Email Integration (via IMAP)
- Notifications
- Enhancements to the user interface

[Read the complete article](#)

[Guest Accounts Allow Off-Campus Collaboration](#)

Berkeley faculty and staff: now you can use Research Hub to collaborate with off-campus colleagues. [Learn more](#)

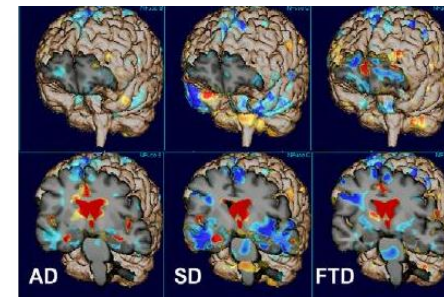
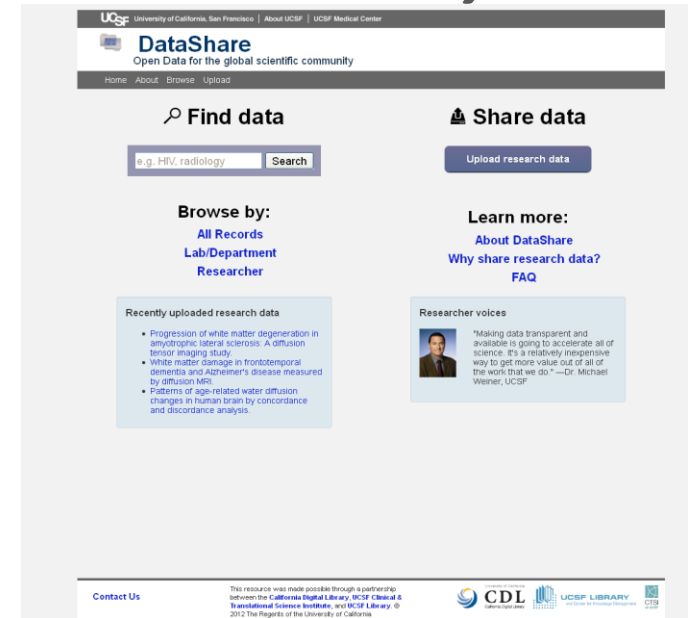
Merritt example: DataShare bright archive / direct discovery



Catalyze widespread intra- and inter-disciplinary sharing of research data

Pilot project in neuro-degenerative disease imaging

Persistent identification and preservation through integration with EZID and Merritt



Merritt example: DataONE— part of a data grid



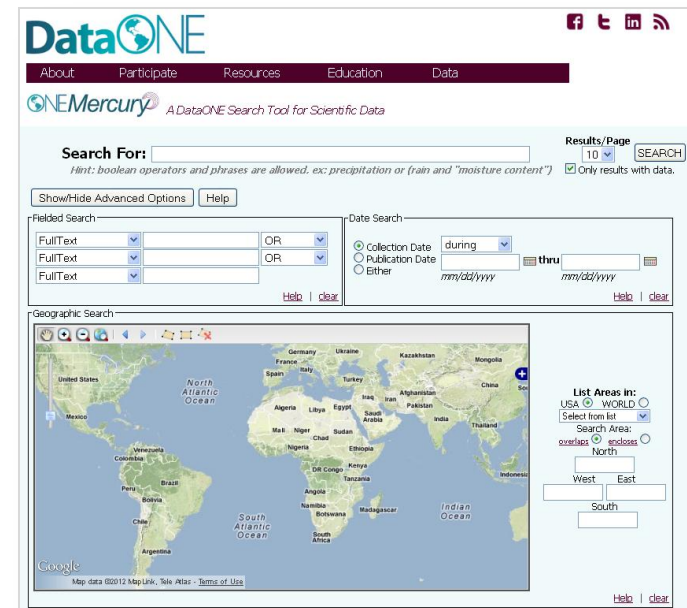
Federated framework and sustainable cyber-infrastructure for innovative, data-intensive environmental science

- Investigator’s toolkit, global replication and discovery, education

A community of engaged stakeholders

- Academic
- Commercial
- Non-profit
- Governmental

 Merritt participates as a DataONE member node



www.dataone.org



ONEShare

Data Repository for **Anyone** | **Anywhere**

Post as CSV

 Eaffinis_2008experiments.xlsx



Submit your data to a repository for long-term archiving. If you do not have an existing relationship with any repository, you will be prompted to proceed.

Repository*

- ✓ ONEShare
- Practice Repository

[User Agreement](#)

I Accept

Dryad
eBird
USGS
Figshare
...

Service 5: Web Archiving Service (WAS)



- ARCHIVE institution websites
- BUILD collections for research
- CAPTURE political and social events
- SAVE at-risk government websites

A screenshot of the Los Angeles Local Government Web Archive website. The page features a header with various logos, a navigation menu, and a main content area with a search bar and a description of the archive. Below the main content, there is a section for "California Political Blogs" and a "Further Information" section. The bottom of the screenshot shows a "Refine site list" section with a search bar and a list of topics.

Primary Functions

1. Capture
2. Manage
3. Preserve
4. Publish



Cost modeling

The long-term success of Merritt and its surrounding services is dependent on their sustainability: technically, administratively, and financially

- Goals
 - *Partial recovery of UC3 service costs*
 - *Flexibility in billing models to ensure uninterrupted service*
- Two options
 - *Pay-as-you-go* (annual payments)
 - *Paid-up* (one-time, up-front payment)



<http://wiki.ucop.edu/display/Curation/Cost+Modeling>

Pivot and work differently

- ✓ Engage with the research community early in the lifecycle – **build community**
- ✓ Prioritize the use of scarce resources to support emerging initiatives – **be entrepreneurial**
- ✓ Start small: provide simple solutions that build up to solve more complex problems – **develop sustainability solutions**
- ✓ Deploy simple and flexible infrastructure and services that can be used in diverse ways – **enable applications to be used in diverse ways**

Develop partnerships and collaborate now more than ever!

For more information

UC Curation Center

<http://www.cdlib.org/uc3>

uc3@ucop.edu

Stephen Abrams

Mark Reyes

Patricia Cruse

Abhishek Salve

Scott Fisher

Joan Starr

Erik Hetzner

Rosalie Lack

Greg Janée

Carly Strasser

John Kunze

Marisa Strong

Margaret Low

Adrian Turner

David Loy

Perry Willett

