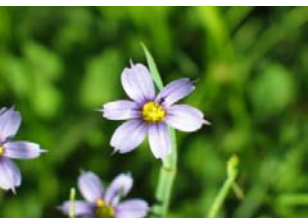




Libraries as Partners in Research: the UC Curation Center's Tools and Services

Patricia Cruse

*University of California Curation Center
California Digital Library*



data intensive research

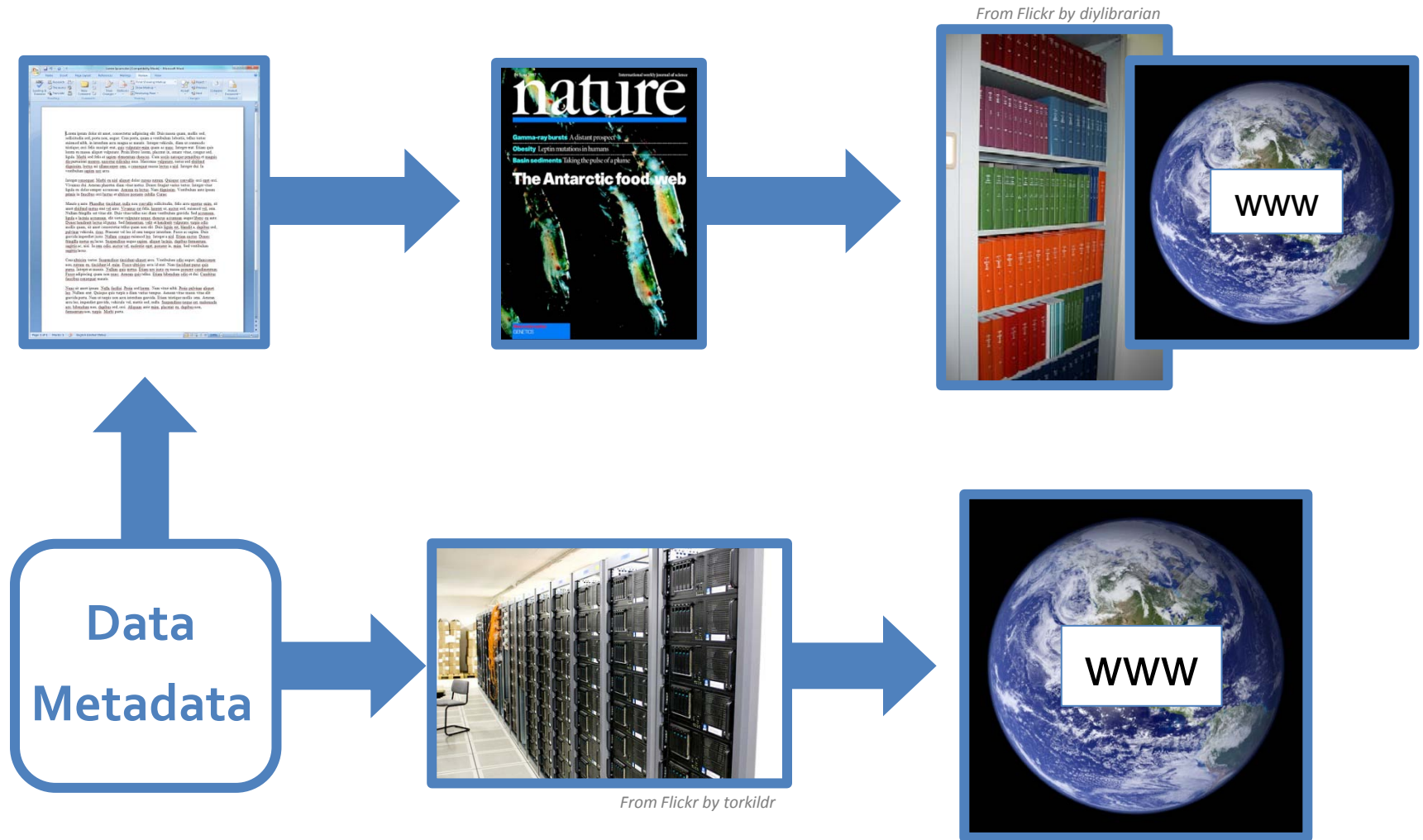
- Ever increasing number, size, and diversity of content
- Ever increasing diversity of partners, and stakeholders
- Data are the new oil: “the future belongs to those that turn data into products” Mike Loukides



Hypercities UCLA







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.

Are data included in open access?

Who owns my data?

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?



Challenge for Researchers



Role of libraries

apply library knowledge and expertise to data challenges: **manage, organize, describe, disseminate, preserve information**

- Neutral service provider: work across the entire institution
- Intellectual Property experts: dealt with copyright, can translate to data
- Ability to advise on preserving research outputs
- Knowledge to advise on data management and curation
- Knowledge on complying with funder mandates, including open access

Resulting service landscape

Create, edit, share, and save data
management plans



Create, share, archive,
publish data



Create and manage persistent
identifiers



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



An infrastructure to publish and get credit
for sharing research data



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



DMPTool

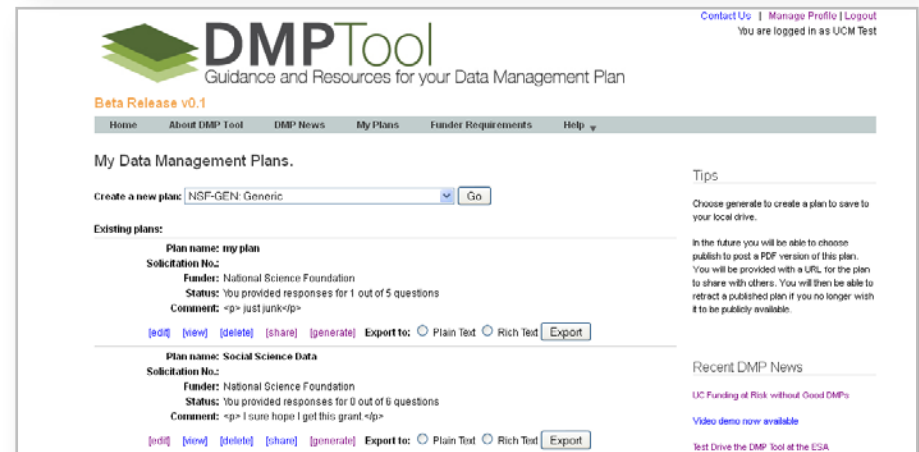
Meeting funding agencies data management plan requirements

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

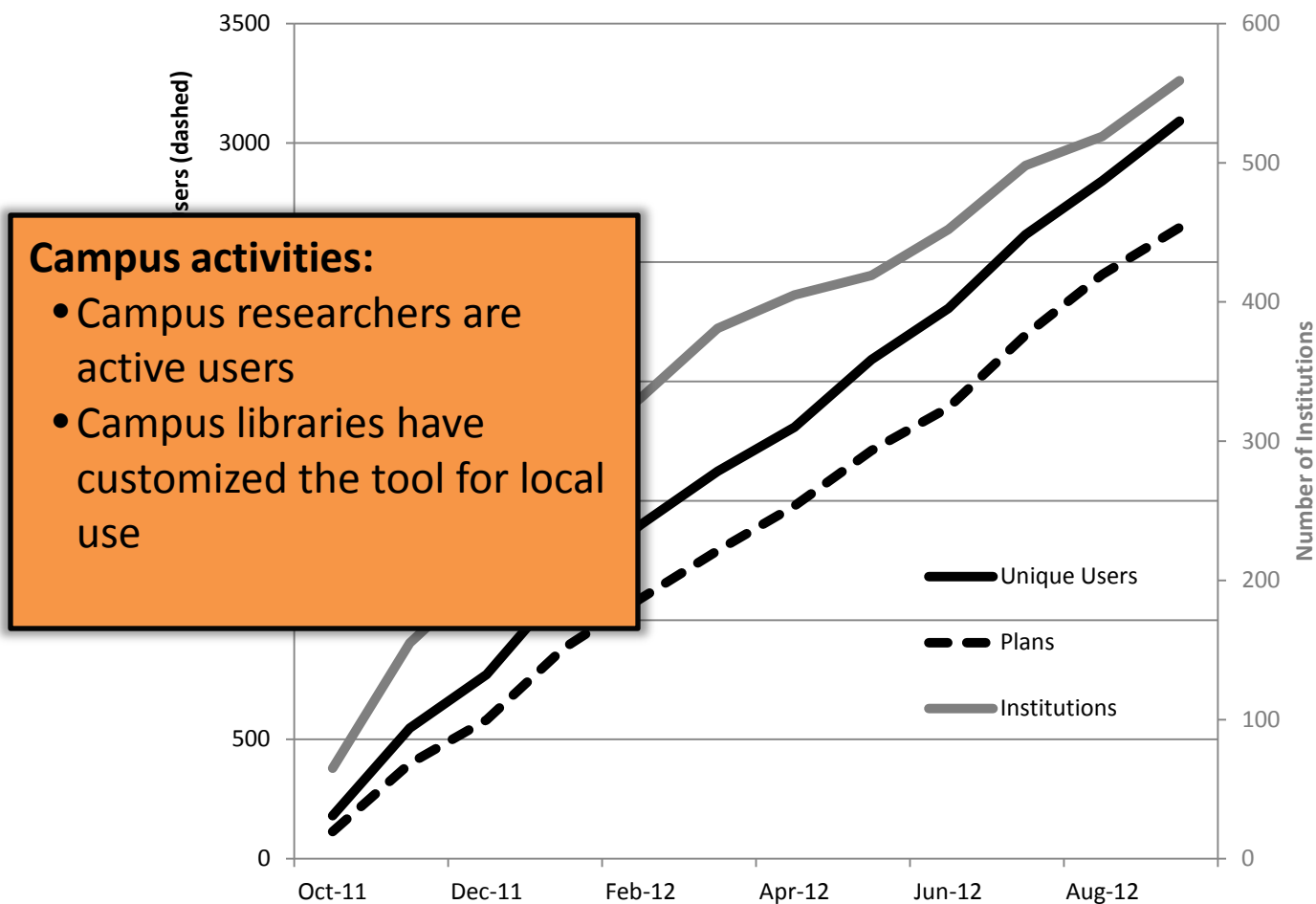


Primary Functions

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



DMTool usage

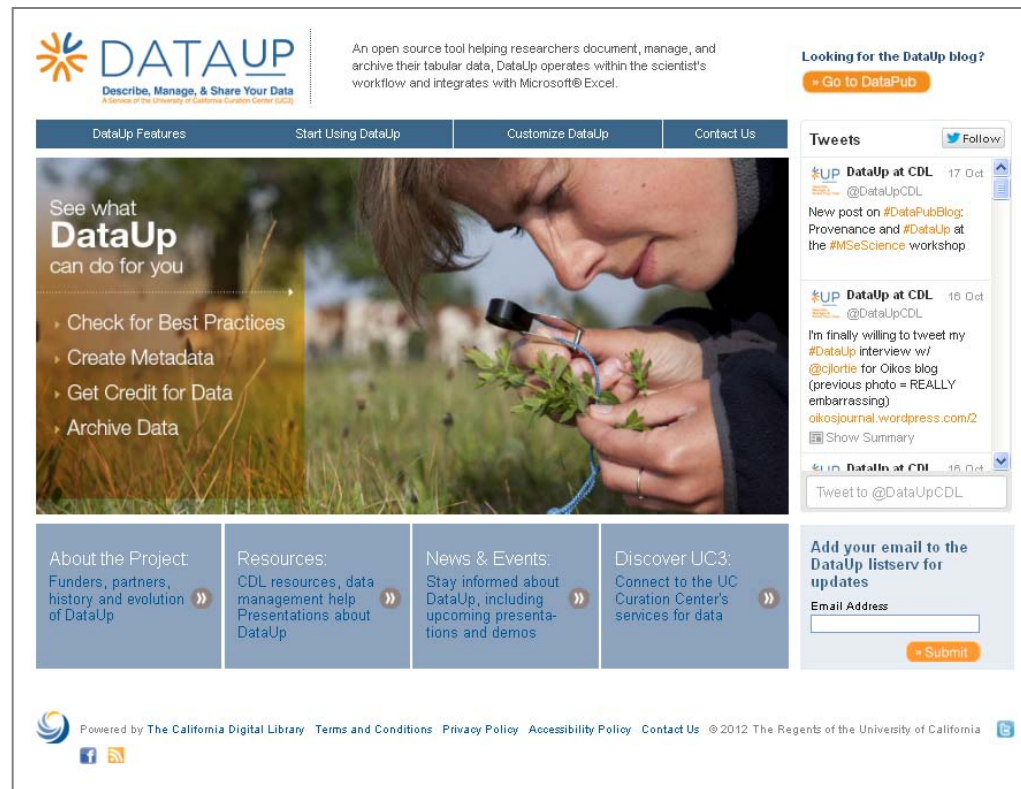


Create, share, archive, connect, publish data

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



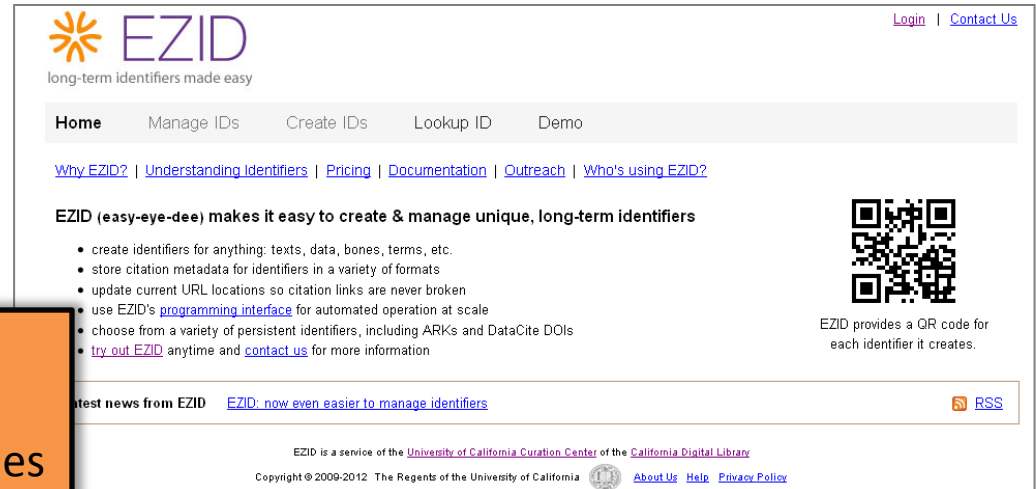
The screenshot shows the DataUp website interface. At the top, the DataUp logo is displayed with the tagline "Describe, Manage, & Share Your Data". Below the logo, a navigation bar includes links for "DataUp Features", "Start Using DataUp", "Customize DataUp", and "Contact Us". A main banner image shows a person using a magnifying glass on a plant, with text overlay stating "See what DataUp can do for you". A list of features is provided: "Check for Best Practices", "Create Metadata", "Get Credit for Data", and "Archive Data". To the right, a sidebar section titled "Looking for the DataUp blog?" includes a "Go to DataPub" button and a "Tweets" section with recent tweets from @DataUpCDL. At the bottom, a footer contains links for "About the Project", "Resources", "News & Events", and "Discover UC3", along with a contact form for email updates. The footer also mentions "Powered by The California Digital Library" and includes social media icons for Facebook and Twitter.

EZID: Long term identifiers made easy

- Precise identification of a dataset (DOI or ARK)
- Credit to data producers and data publishers

- A link from the traditional literature

- Ex for (W
- Campus library activities:**
 - Nine out of ten campus libraries subscribe to the service
 - Campus libraries monitor how to expand / index



1. Create persistent identifiers
2. Manage identifiers (and associated metadata) over time
3. Resolve identifiers

Merritt Repository



- Curation repository open to the UC community and beyond
- Discipline / content agnostic
- New kind of cost model: *pay as you go* or *pay once and store for X*

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 Campus Needs

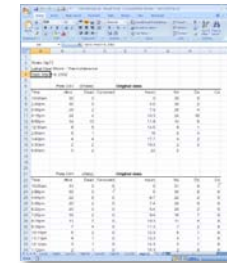
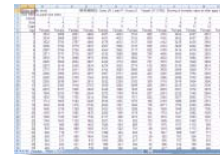
- Dark archive for important digital assets



- Bright archive with direct

Campus library activities:

- Use the service to archive library content
- Reach out to researchers
- Help grow the service



for
ery
ent

systems



- Integration with distributed data grids





DataShare

Open Data for the global scientific community

[Home](#) [About](#) [Browse](#) [Upload](#)

Find data

Browse by:

[All Records](#)

[Lab/Department](#)

[Researcher](#)

Recently uploaded research data

- [Progression of white matter degeneration in amyotrophic lateral sclerosis: A diffusion tensor imaging study.](#)
- [White matter damage in frontotemporal dementia and Alzheimer's disease measured by diffusion MRI.](#)
- [Patterns of age-related water diffusion changes in human brain by concordance and discordance analysis.](#)

Share data

[Upload research data](#)

Learn more:

[About DataShare](#)

[Why share research data?](#)

[FAQ](#)

Researcher voices



"Making data transparent and available is going to accelerate all of science. It's a relatively inexpensive way to get more value out of all of the work that we do." —Dr. Michael Weiner, UCSF

Vision for a “data paper”

- Wrap the unfamiliar in a familiar façade
- Minimally, a cover sheet and a set of links to archived artifacts
- Cover sheet contains familiar elements: title, date, authors, abstract, identifiers
- Just enough metadata to permit basic exposure to and discovery
 - Indexing by services such as Web of Science, Google Scholar
 - Instilling confidence in the identifier’s stability

Multi-decade, spatially explicit population studies of canopy dynamics in Michigan old-growth forests

Data Paper. 2009. doi:10.5060/D2E090/251

Kerry D. Woods

Natural Sciences, Bennington College, Bennington, Vermont 05201 USA

Abstract

Established in 1935, a regular grid of 256 permanent plots includes about 20% of a 100-ha old-growth forest at the Dukes Research Natural Area in northern Michigan, USA. Woody stems have been remeasured 3–7 times providing extensive quantitative records of population and community dynamics over periods of up to 72 years. Woody stems in upland hemlock–northern hardwood stands, about half of the study plots, have been mapped and individually tracked since about 1990. Remaining plots are in swampy stands dominated by *Fraxinus nigra* and *Thuja occidentalis*. Detailed, long-term demographic data for late-successional forests are rare in general; this data set is both of exceptional duration and unusual in spatial intensity and detail. Because sample plots are in a regular array over the stand, they can support analyses of spatiotemporal pattern at various scales. A major wind disturbance in 2002 provides a unique opportunity to compare disturbance response to baseline dynamics. Several publications based on this data set have already provided new insights into late-successional processes, but general availability of the data set with meta data should permit a range of further comparative and integrative analyses. The study is ongoing, and new measurements will be added to the archived data set. Several ancillary data sets are available from the author.

Key words: *Acer saccharum*; *Betula alleghaniensis*; *Fagus grandifolia*; *Fraxinus nigra*; *long-term studies*; *northern hardwood forest*; *old-growth forest*; *permanent plots*; *succession*; *Thuja occidentalis*; *tree mapping*; *Tsuga canadensis*.

Data Files

Files are ASCII text, tab-delimited. No compression schemes were used.

[all_plots_1935_1948.txt](#) -- data for all stems measured in 1935 and 1948.

[all_plots_1974-1980.txt](#) -- data for all stems measured in 1974 through 1980.

[upland_plots_89-07.txt](#) -- data for upland plots mapped and measured two or more times, 1989 through 2007.

[swamp_all_modern.txt](#) -- data for wetland plots censused from 1992 through 2007.

[species_codes.txt](#) -- four-letter codes and full names for all species.

[sampling_history.txt](#) -- table summarizing sampling history for all plots.

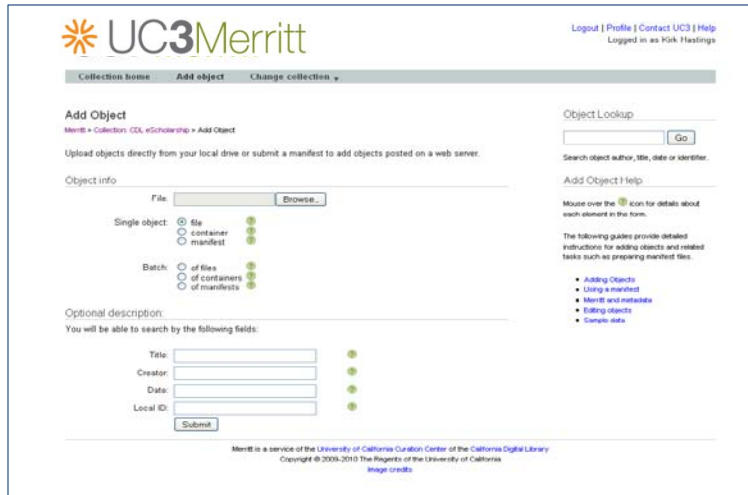
Data Publishing at the CDL

UC Curation Center

- Merritt Curation repository
- EZID: Persistent id management and resolution (ARKs, DOIs, et al.)

Publishing Services Program

- Online journals, with peer review
- Scholarly communication: grey literature to post-prints
- Search and display tools (XTF)



The screenshot shows the 'Add Object' page of the UC3 Merritt repository. The page has a header with the UC3 Merritt logo and navigation links. The main content area is divided into two columns. The left column contains the 'Add Object' form, which includes a 'File' upload section with a 'Browse...' button, a 'Batch' section for uploading multiple files, and an 'Optional description' section with fields for Title, Creator, Date, and Local ID. The right column contains an 'Object Lookup' section with a search box and a 'Go' button, and an 'Add Object Help' section with links to various guides. The footer of the page states that Merritt is a service of the University of California Curation Center of the California Digital Library.



The screenshot shows the eScholarship website, which is a digital library and publishing platform. The page features a header with the eScholarship logo and navigation links. The main content area is divided into two columns. The left column contains a 'Journal Info' section for 'California Agriculture', which includes links to the journal's website, about page, contact information, and submission guidelines. The right column contains a 'Current Issue' section for 'Volume 65, Issue 1, 2011', which lists several articles with their titles and authors. The footer of the page includes a 'My Account' link and a statement that eScholarship is powered by the California Digital Library and The Berkeley Electronic Press.



University of California Curation Center Web Archiving Service

The screenshot displays the Web Archiving Service interface. At the top, there's a navigation bar with links for Home, Help, and Contact WAS. Below this, a status bar indicates the user is logged in as 'tseneca' and shows the current project as 'California State Government: .ca.gov Web Archive'. The main content area is divided into four tabs: Sites, Captures, Administration, and Public Access. The 'Sites' tab is active, showing a list of sites to be archived. The '1. Create Site' section includes a form for entering site details. The '2. Capture Site' section shows a list of sites and a 'Recent Capture Activity' box. The '3. View Captures' section shows a list of captured sites. The '4. Public Access' section shows a preview of the 'African Politics Web Archive' page, which includes a search bar and a list of archived sites.

53 public archives
120+ archives total
58K crawls
7,500 + sites
600 million + URLs
60+ TB
24 institutions

What are people using WAS for?

Archiving at-risk websites and publications

Archiving their own university domains

Building web archives to complement library collections

Documenting web coverage of significant events



Los Angeles Local Government Web Archive

UCLA Library

[Home](#) [About](#) [Site List](#) [Search](#) [Help](#) [Contact Us](#)

Des...

Los A
Unit
to the
agenc
critic

The v
esped
counc
or as
agenc
have

more

California Political Blogs

by Institute of Governmental Studies

Archives

@ the Bentley Historical Library

edu (734)764-3482 | 150 Beal Ave. Ann Arbor, MI 48109

which are
nion, and other ad
archive preserves

Campus library activities:

- Build collections: locally and collaboratively
- Provide service to campus community

Go Clear

Site list by topic:

- Administration
- Architecture & Urban Planning
- Athletics
- College of Engineering
- College of Lit. Science & Arts
- College of Pharmacy
- Faculty
- Law School
- Medical School
- Museums, Libraries & Culture
- Music, Theatre & Dance
- News & Events
- Rackham Graduate School
- Ross Business School

American Library Association Chapter Web Archives (University of Michigan) [Show Info](#)

Arts on Earth Web Archives (University of Michigan) [Show Info](#)

CAMILEON Project Web Archives (University of Michigan) [Show Info](#)

Center for Cognitive Architecture Web Archives (University of Michigan) [Show Info](#)

Center for Information Technology Integration Web Archives (University of Michigan) [Show Info](#)

Collaboratory for Research on Electronic Work Web Archives (University of Michigan) [Show Info](#)

Community Information Corps Web Archives (University of Michigan) [Show Info](#)

ure. This delay is in

al Library. This
with the support of
e was made
ervation Program

New kind of cost model

- Understand *costs* in order to plan for and implement *sustainable* preservation services
- Investigate *paid-up* pricing in order to address
 - Boom-or-bust budget cycles
 - Fixed-term, grant funded projects



Source: www.sharedidiz.com/

Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31		

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

