



*From Preservation to Curation:
extending boundaries, creating new services,
engaging new users*

*Patricia Cruse / Stephen Abrams
University of California Curation Center
California Digital Library
NDIIPP / NDSA Partner Meeting
July 19-21, 2011*



Our environment circa 2002-2008

Focus on preservation

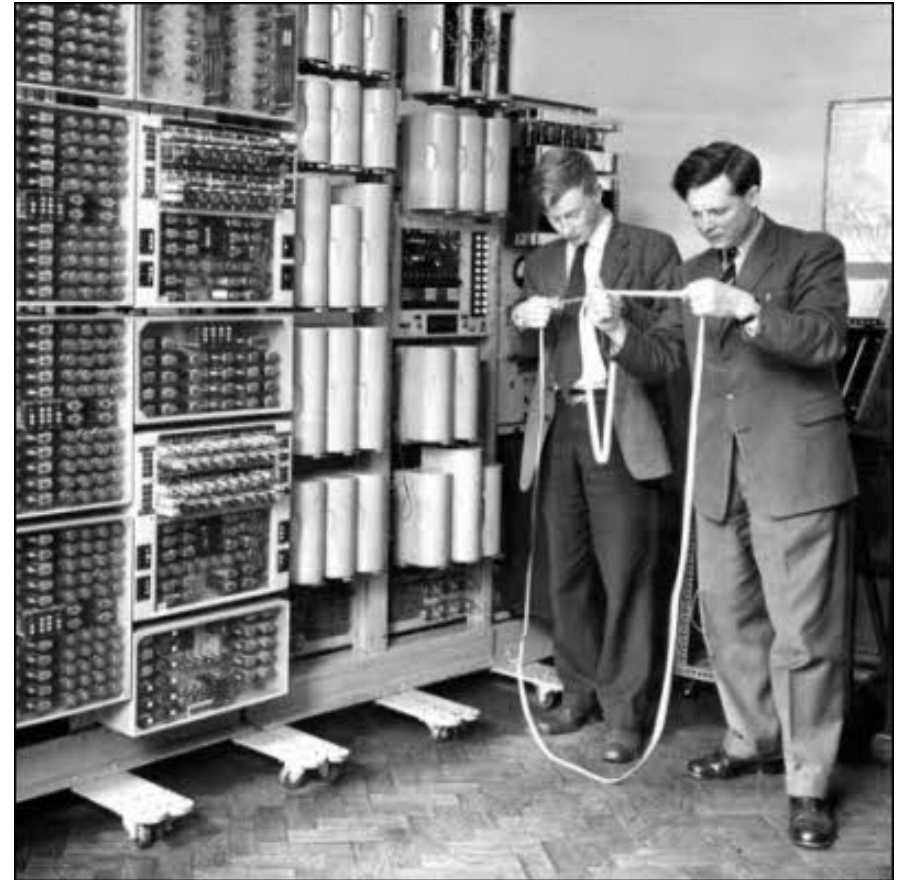
Stakeholders: memory
organizations

Infrastructure: static

Services: hosted

Content: museum and
library

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
- Users – changing expectations and needs



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?

How can I share my work with my colleagues?

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

How can I provide access to my work?

What is "metadata"?

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?



Four questions or imperatives?

- How can we best respond organizationally?
- How does our technical landscape change?
- What is the value of our services to our diverse community of users?
- How can we build (or reach) new communities?

University of California Curation Center

Creative partnership between the CDL, the 10 UC campuses, individuals and peer institutions

- A community of shared concern and practice
- A channel to pool and distribute diverse experience, expertise, and resources
- Robust, innovative, and cost-effective solutions to counteract inevitable disruptive change



UC Curation Center's environment today

Organization & Stakeholders:

UC libraries, UC community, and beyond

Focus on curation and entire information lifecycle

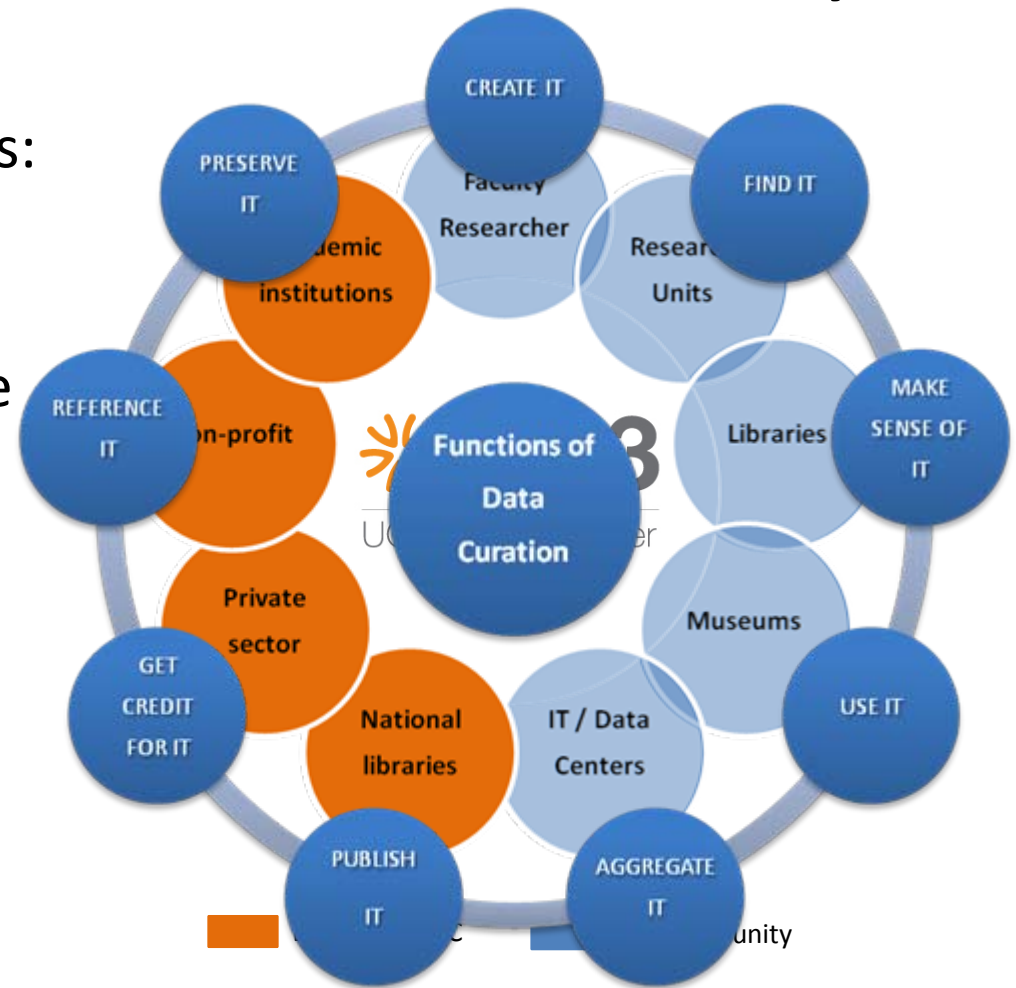
Technology and

Infrastructure: simple, flexible, adaptable

Services: diverse

Content: agnostic

Sustainability: a must



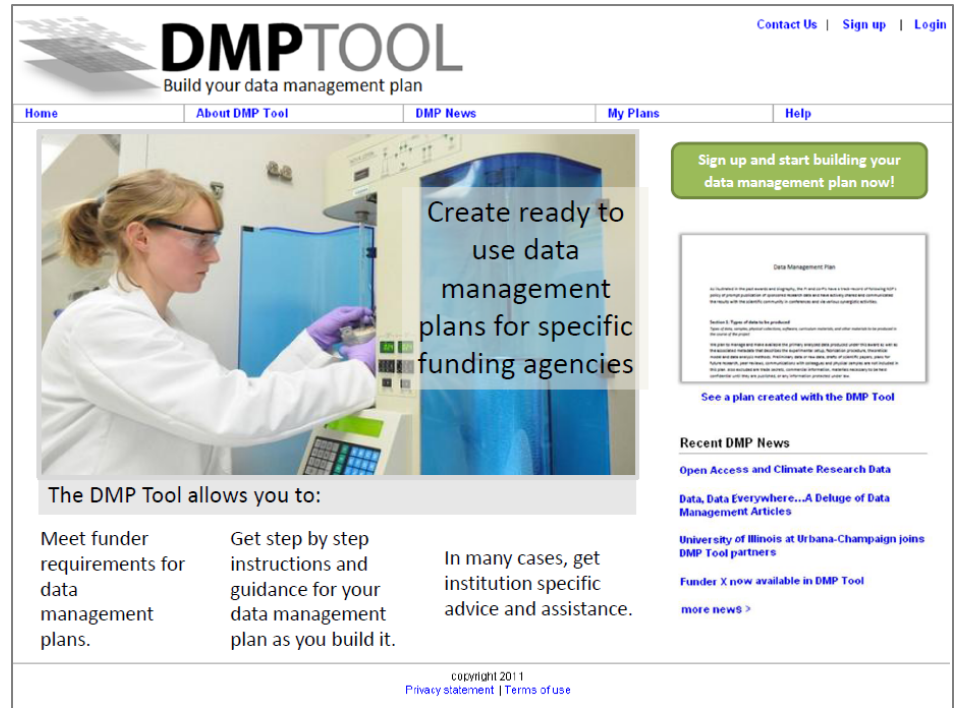
Data Management Planning (DMP) Tool

Funding agencies requiring a DMP

1. connect researchers with resources
2. streamline the process to produce a credible and high-quality plan for managing data

Eight institutions coming together

Tool will have multiple phases



The screenshot shows the homepage of the DMP Tool website. At the top left is the logo for the DMP Tool, which includes a keyboard icon and the text "DMP TOOL Build your data management plan". To the right of the logo are links for "Contact Us", "Sign up", and "Login". Below the logo is a navigation bar with links for "Home", "About DMP Tool", "DMP News", "My Plans", and "Help". The main content area features a large image of a scientist in a lab coat working with a piece of equipment. Overlaid on this image is the text "Create ready to use data management plans for specific funding agencies". To the right of the image is a green button that says "Sign up and start building your data management plan now!". Below the image is a section titled "The DMP Tool allows you to:" followed by three columns of text: "Meet funder requirements for data management plans.", "Get step by step instructions and guidance for your data management plan as you build it.", and "In many cases, get institution specific advice and assistance." To the right of this section is a "Recent DMP News" section with several links to news articles. At the bottom of the page, there is a copyright notice for 2011 and links for "Privacy statement" and "Terms of use".

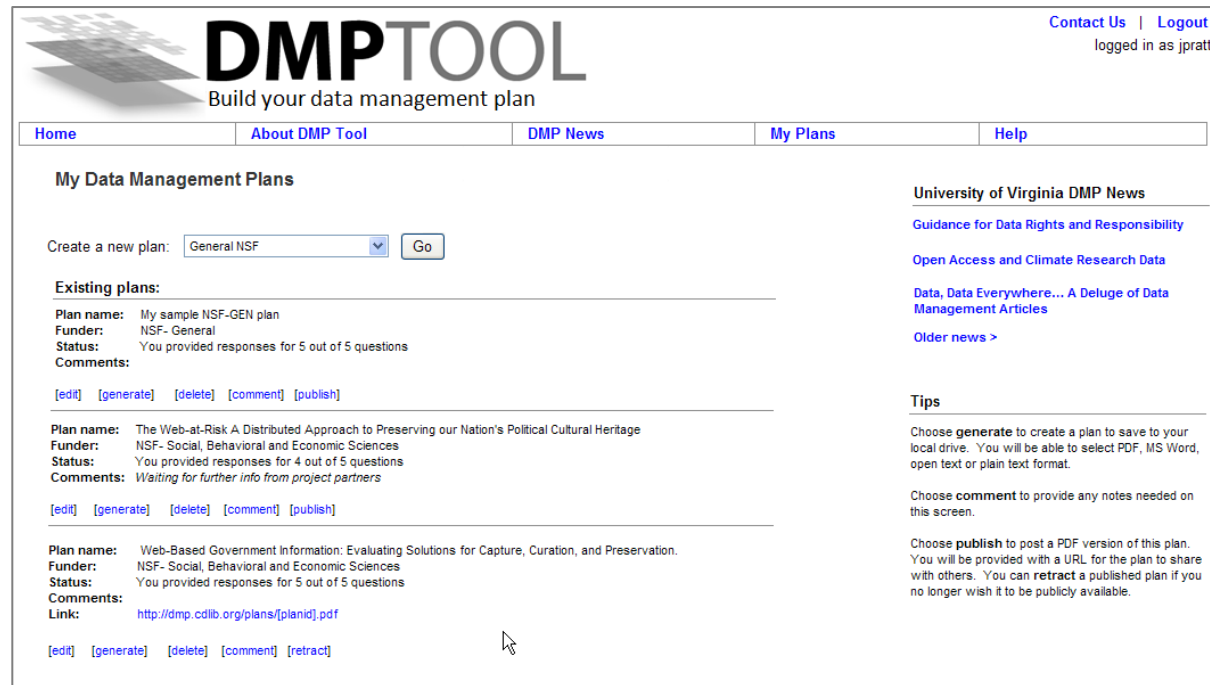
DMP Tool Out-of-the-Box

1. For all users

- Step-by-step ‘wizard’ for generating data management plans
- General guidance for each section: help text and resources relevant to all
- Save a plan as PDF, MS Word, plain text or generate a link to a PDF version of the finished plan

2. For DMP Tool Partners

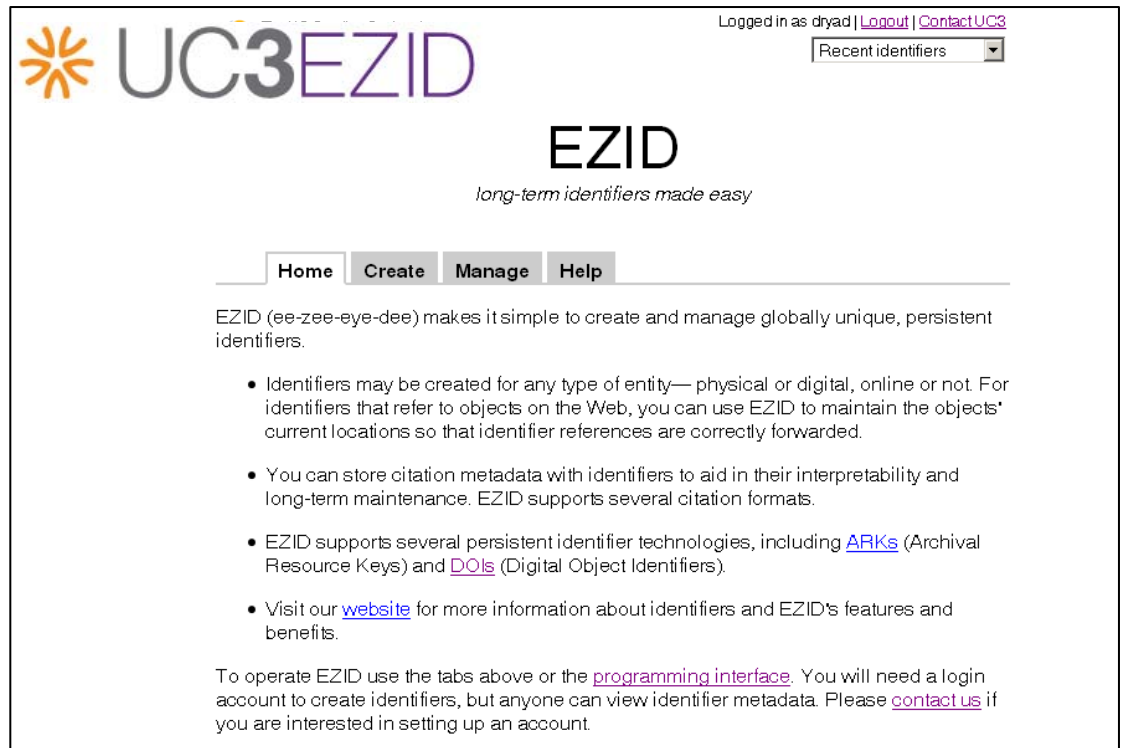
- Customized links to resources available to all institution’s researchers



The screenshot shows the DMP Tool website interface. At the top, there is a navigation bar with links for Home, About DMP Tool, DMP News, My Plans, and Help. The main content area is titled "My Data Management Plans" and features a "Create a new plan" section with a dropdown menu set to "General NSF" and a "Go" button. Below this, there are three existing plans listed, each with a "Plan name", "Funder", "Status", and "Comments" field, and a set of action links: [edit], [generate], [delete], [comment], and [publish]. The first plan is "My sample NSF-GEN plan" with a status of "You provided responses for 5 out of 5 questions". The second plan is "The Web-at-Risk A Distributed Approach to Preserving our Nation's Political Cultural Heritage" with a status of "You provided responses for 4 out of 5 questions". The third plan is "Web-Based Government Information: Evaluating Solutions for Capture, Curation, and Preservation." with a status of "You provided responses for 5 out of 5 questions" and a link to a PDF version. On the right side of the page, there is a "University of Virginia DMP News" section with links for "Guidance for Data Rights and Responsibility", "Open Access and Climate Research Data", "Data, Data Everywhere... A Deluge of Data Management Articles", and "Older news >". Below this is a "Tips" section with instructions on how to use the "generate", "comment", and "publish" buttons.

EZID: long-term identifiers made easy

take control of the management and distribution of your research, share and get credit for it, and build your reputation through its collection and documentation



The screenshot shows the UC3 EZID website. At the top left is the UC3 logo. To the right, it says "Logged in as dryad | [Logout](#) | [Contact UC3](#)". Below that is a dropdown menu labeled "Recent identifiers". The main heading is "EZID" with the tagline "long-term identifiers made easy". A navigation bar contains tabs for "Home", "Create", "Manage", and "Help". The "Manage" tab is selected. The main content area explains that EZID (ee-zee-eye-dee) makes it simple to create and manage globally unique, persistent identifiers. It lists four bullet points: 1. Identifiers may be created for any type of entity—physical or digital, online or not. For identifiers that refer to objects on the Web, you can use EZID to maintain the objects' current locations so that identifier references are correctly forwarded. 2. You can store citation metadata with identifiers to aid in their interpretability and long-term maintenance. EZID supports several citation formats. 3. EZID supports several persistent identifier technologies, including [ARKeys](#) (Archival Resource Keys) and [DOIs](#) (Digital Object Identifiers). 4. Visit our [website](#) for more information about identifiers and EZID's features and benefits. At the bottom, it states: "To operate EZID use the tabs above or the [programming interface](#). You will need a login account to create identifiers, but anyone can view identifier metadata. Please [contact us](#) if you are interested in setting up an account."

Primary Functions

1. Create persistent identifiers
2. Manage identifiers over time
3. Manage associated metadata over time

EZID supports a wish list for data as a key component of scholarly communication

Supporting researchers

- Precise identification of a dataset
- Credit to data producers and data publishers
- A link from the traditional literature to the data
- Research metrics for datasets

Supporting a community

Business model

- Tiered pricing structure for UC, non-UC, for profit
- Revenue supports operations and development
- Range of customers: government agencies, research centers, institutions, for-profit

Working with publishers to expose data as publication

UC3 Merritt



Service overview

Open to the UC community & beyond

Discipline / content agnostic

Service delivery: hosted or local
deployed

Easy to use UI or API

Primary Functions

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

Merritt's diverse service offering to the community

Merritt's service offering

- 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
- Local deployments

Supporting the community

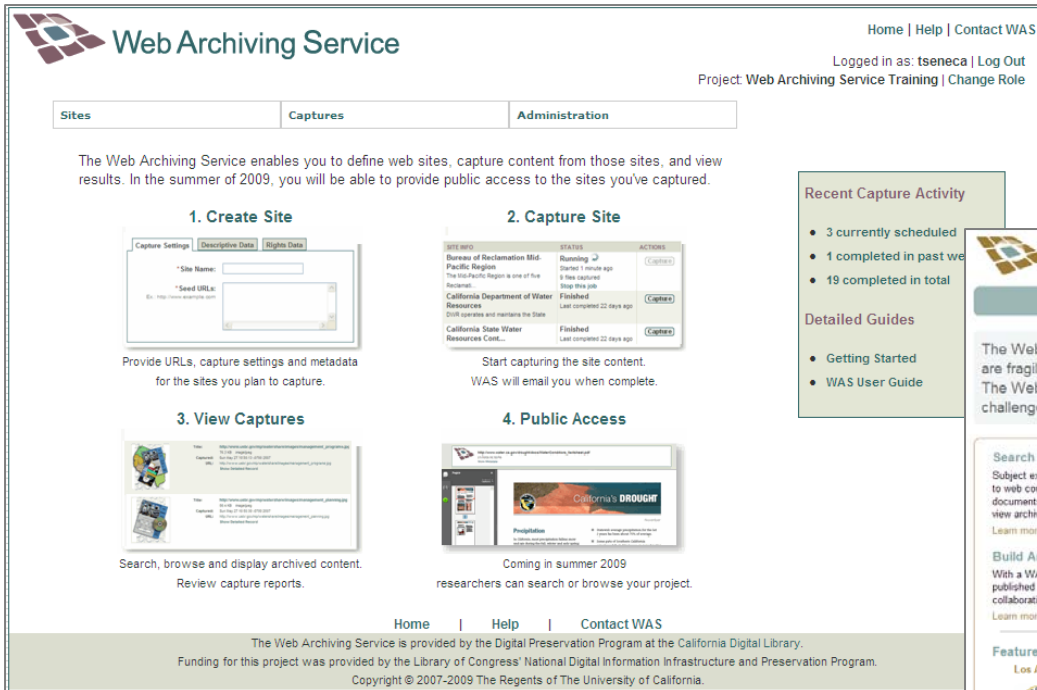
Business model

- Pricing structure for UC, non-UC, and for-profit
 - Pay as you go
 - Pay once store forever
- Revenue supports operation and development
- Range of customers: government agencies, research centers, institutions, for-profit



Web Archiving Service

Capture today's web, build tomorrow's archive



Home | Help | Contact WAS

Logged in as: tseneca | Log Out
Project: Web Archiving Service Training | Change Role

Sites | **Captures** | **Administration**

The Web Archiving Service enables you to define web sites, capture content from those sites, and view results. In the summer of 2009, you will be able to provide public access to the sites you've captured.

- 1. Create Site**
Provide URLs, capture settings and metadata for the sites you plan to capture.
- 2. Capture Site**
Start capturing the site content. WAS will email you when complete.
- 3. View Captures**
Search, browse and display archived content. Review capture reports.
- 4. Public Access**
Coming in summer 2009 researchers can search or browse your project.

Home | Help | Contact WAS

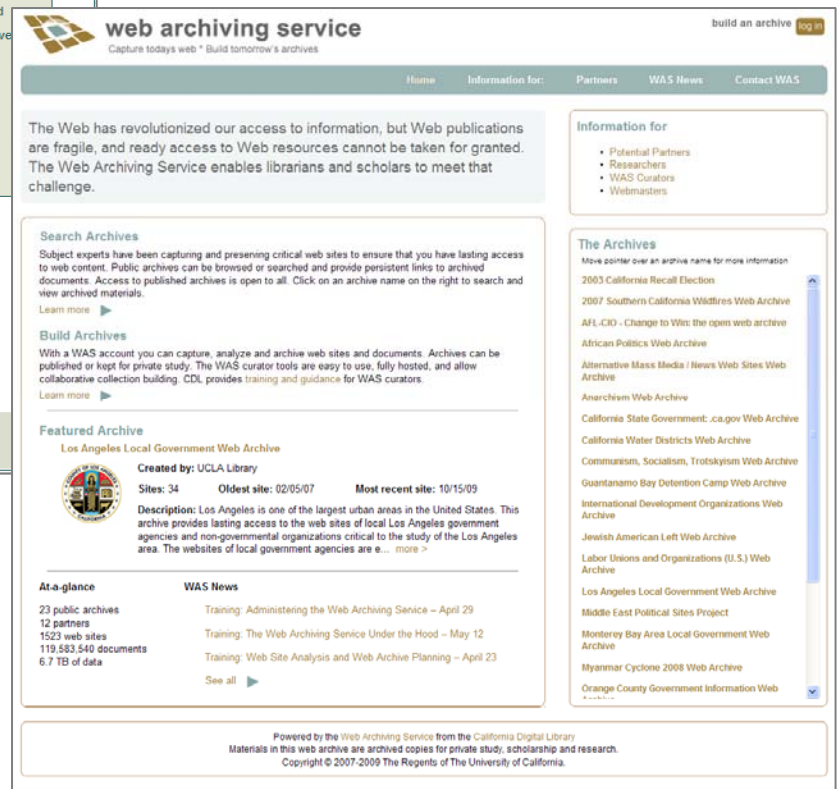
The Web Archiving Service is provided by the Digital Preservation Program at the California Digital Library. Funding for this project was provided by the Library of Congress' National Digital Information Infrastructure and Preservation Program. Copyright © 2007-2009 The Regents of The University of California.

Recent Capture Activity

- 3 currently scheduled
- 1 completed in past week
- 19 completed in total

Detailed Guides

- Getting Started
- WAS User Guide



web archiving service
Capture today's web * Build tomorrow's archives

Home | Information for: | Partners | WAS News | Contact WAS

build an archive [log in](#)

The Web has revolutionized our access to information, but Web publications are fragile, and ready access to Web resources cannot be taken for granted. The Web Archiving Service enables librarians and scholars to meet that challenge.

Information for

- Potential Partners
- Researchers
- WAS Curators
- Webmasters

The Archives

Move pointer over an archive name for more information

- 2003 California Recall Election
- 2007 Southern California Wildfires Web Archive
- AFL-CIO - Change to Win: the open web archive
- African Politics Web Archive
- Alternative Mass Media / News Web Sites Web Archive
- Anarchism Web Archive
- California State Government: .ca.gov Web Archive
- California Water Districts Web Archive
- Communism, Socialism, Trotskyism Web Archive
- Guantanamo Bay Detention Camp Web Archive
- International Development Organizations Web Archive
- Jewish American Left Web Archive
- Labor Unions and Organizations (U.S.) Web Archive
- Los Angeles Local Government Web Archive
- Middle East Political Sites Project
- Montgomery Bay Area Local Government Web Archive
- Myanmar Cyclone 2008 Web Archive
- Orange County Government Information Web Archive

Search Archives

Subject experts have been capturing and preserving critical web sites to ensure that you have lasting access to web content. Public archives can be browsed or searched and provide persistent links to archived documents. Access to published archives is open to all. Click on an archive name on the right to search and view archived materials.

Build Archives

With a WAS account you can capture, analyze and archive web sites and documents. Archives can be published or kept for private study. The WAS curator tools are easy to use, fully hosted, and allow collaborative collection building. CDL provides training and guidance for WAS curators.

Featured Archive

Los Angeles Local Government Web Archive

Created by: UCLA Library

Sites: 34 | Oldest site: 02/05/07 | Most recent site: 10/15/09

Description: Los Angeles is one of the largest urban areas in the United States. This archive provides lasting access to the web sites of local Los Angeles government agencies and non-governmental organizations critical to the study of the Los Angeles area. The websites of local government agencies are e... more >

At-a-glance

23 public archives
12 partners
1523 web sites
119,883,540 documents
6.7 TB of data

WAS News

Training: Administering the Web Archiving Service - April 29
Training: The Web Archiving Service Under the Hood - May 12
Training: Web Site Analysis and Web Archive Planning - April 23

See all >

Powered by the Web Archiving Service from the California Digital Library
Materials in this web archive are archived copies for private study, scholarship and research.
Copyright © 2007-2009 The Regents of The University of California.

Primary Functions

1. Collect web published content
2. Manage content
3. Publish content for public access

WAS service overview



web archiving service
Capture today's web * Build tomorrow's archives

Home Information for: Partners WAS News Contact WAS

Web archives from *Stanford University Library*

African Politics Web Archive



Created by: Stanford University Library
Sites: 29 **Oldest site:** 12/05/08 **Most recent site:** 12/09/09

Description: The African Politics Web Archive seeks to preserve web sites for future generations as traditional libraries preserve books, journals, newspapers and other materials. Presidential election web sites generally last only a few months on the web and then disappear. Web sites archived appear in th... [more >](#)

Middle East Political Sites Project



Created by: Stanford University Library
Sites: 145 **Oldest site:** 01/31/07 **Most recent site:** 12/10/09

Description: The project intends to capture and archive the web sites by and about some of the more important political organizations and related NGOs from the countries of the Middle East.

Business model in place

- Pricing structure for UC, non-UC, and for profit
 - Service fee and storage used
- Revenue supports operations and development
- Range of customers: agencies, research centers, academic institutions, researchers, libraries



Digital Curation for EXCEL (DCXL) Project

Open source MS Excel add-ins

Problem statement

Data are the building blocks of scientific research.

Many scientists use MS Excel to record, manage, view, graph, and manipulate datasets.

Excel's current feature set can be a barrier to sharing, verifying, and preserving

DCXL Outputs

- Requirements (open)
- Open source add-in
 - interoperable,
 - sharable,
 - publishable,
 - archivable
- New community of practice

What an Excel add-in could do

Some preliminary ideas to better publish, share, and archive

- Permit standardized column headers
- Versioning and standard date formats
- Auto-archiving and persistent id assignment
- “Speed bumps” to discourage macros et al.

Participants

- UC3 at the CDL
- UC Campus community
- DataONE
- The broader community
- MS Research
- Gordon and Betty Moore Foundation

Delivery date: Spring 2012

Vision for a “data paper”

- Idea: wrap the unfamiliar in a familiar façade
- A “data paper” minimally consists of a cover sheet and a set of links to archived artifacts
- Cover sheet contains familiar elements: title, date, authors, abstract, and persistent identifier (DOI, ARK, etc.)
- Just enough to permit basic exposure to and discovery
 - Building a basic data citation
 - 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

Keny 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 metadata 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' interface on the UC3 Merritt website. It includes a search bar for object lookup, an 'Add Object Help' section with instructions, and a form for uploading objects. The form has radio buttons for 'Single object' (file, container, manifest) and 'Batch' (files, containers, manifests). There are also input fields for 'Optional description' (Title, Creator, Date, Local ID) and a 'Submit' button. The footer contains copyright information for the University of California.



The screenshot shows the eScholarship website interface. It features a navigation menu with 'Home', 'About', 'Browse', 'Publish', 'Help', and 'My Items (0)'. A search bar is located in the top right. The main content area displays 'California Agriculture' journal information, including a 'Current Issue, Volume 65, Issue 1, 2011' section with a featured article: 'UC Cooperative Extension sensory analysis panel enhances the quality of California olive oil'. The article lists authors: Vossen, Paul, Koceniik-Devarenne, Alexandra. Other articles listed include 'Understanding the seasonal and reproductive biology of olive fruit fly', 'Biological controls investigated to aid management of olive fruit fly in California', 'High temperature affects olive fruit fly populations in California's Central Valley', 'Mediterranean clonal selections evaluated for modern hedgerow olive oil production in Spain', and 'Olive cultivars field-tested in super-high-density system in southern Italy'. The footer includes 'My Account', 'eScholarship is powered by the California Digital Library with The Berkeley Electronic Press', and 'Contact eScholarship'.

Lessons learned (and still learning)

Goal is to work on several fronts to make a complex problems smaller

- Don't circle the wagons
- Stop doing what you can't support
- Outsource and/or use third party components
- Deploy new infrastructure and services that can be used in diverse ways
- Engage with new communities – research community
- Support emerging initiatives
- Collaborate now more than ever!

UC3's special agents at the CDL



Tracy Seneca



Perry Willett



Margaret Low



Mark Reyes



Carly Strasser



Lisa Colvin



Scott Fisher



Marisa Strong



Erik Hetzner



Greg Janee



John Kunze



Stephen Abrams



Trisha Cruse



David Loy