UC3

UC Curation Center

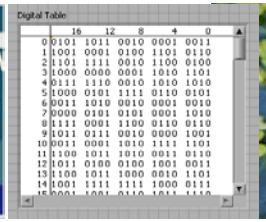
Libraries as Partners in Research: the UC Curation Center's Tools and Services

Patricia Cruse

University of California Curation Center

California Digital Library



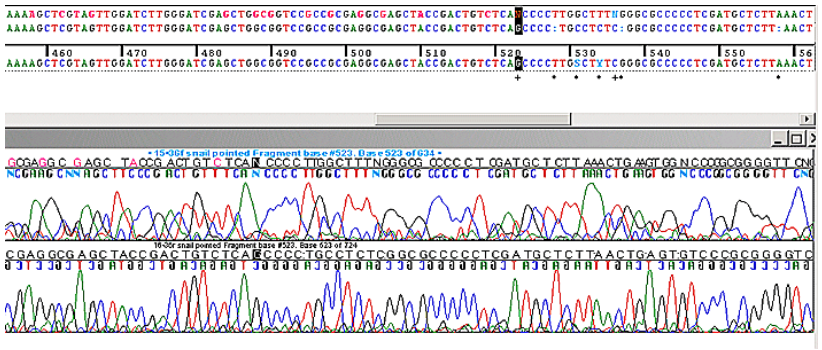


UC3
UC Curation Center

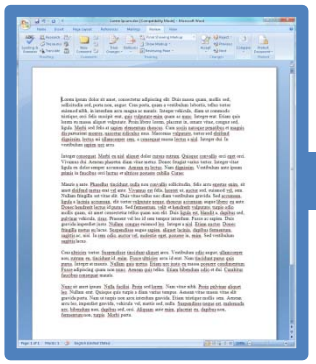
data intensive research

- 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



UC3
UC Curation Center

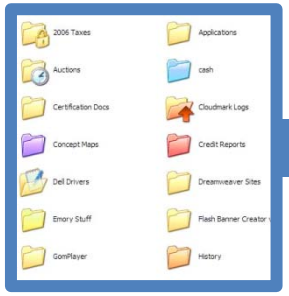


From Flickr by diylibrarian

Data
Metadata

blog.order2disorder.com

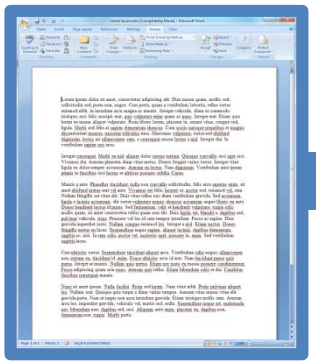
From Flickr by csessums



From Flickr by csessums

Recreated from Klump et al. 2006

UC3
UC Curation Center



From Flickr by diylibrarian

Data

Metadata

From Flickr by torkildr

Recreated from Klump et al. 2006

UC3
UC Curation Center

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?

How can I share my work with my colleagues?

Who owns my data?

How can I provide access to my work?

What are "metadata"?

How can I publish the data associated with my publications?

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

How can I make sure I get credit ?

UC3
UC Curation Center

Challenge for Researchers

UC3
UC Curation Center

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

UC3
UC Curation Center

Resulting service landscape

Create, edit, share, and save data management plans

DMPTool
Guidance and Resources for your Data Management Plan

Create, share, archive, publish data

UC3 DATA UP

Create and manage persistent identifiers

UC3 EZID

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

UC3 Merritt

An infrastructure to publish and get credit for sharing research data

eScholarship
University of California

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

UC3 WAS

UC3
UC Curation Center

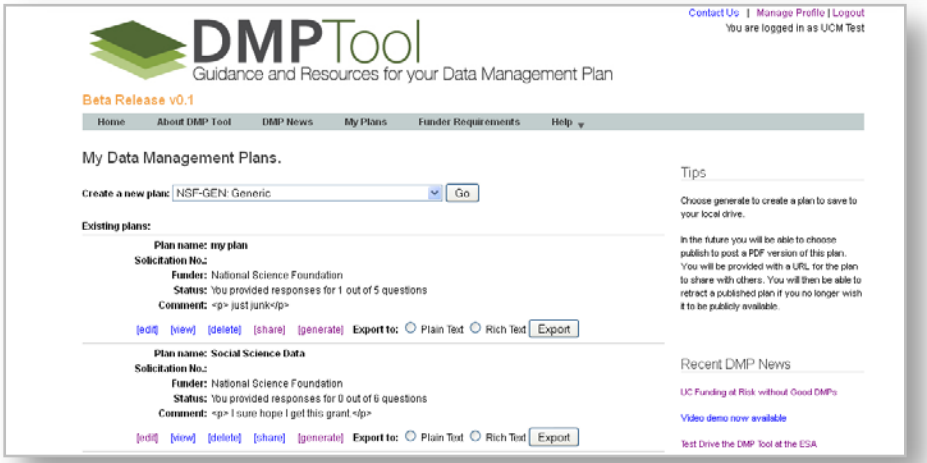
DMPTool

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

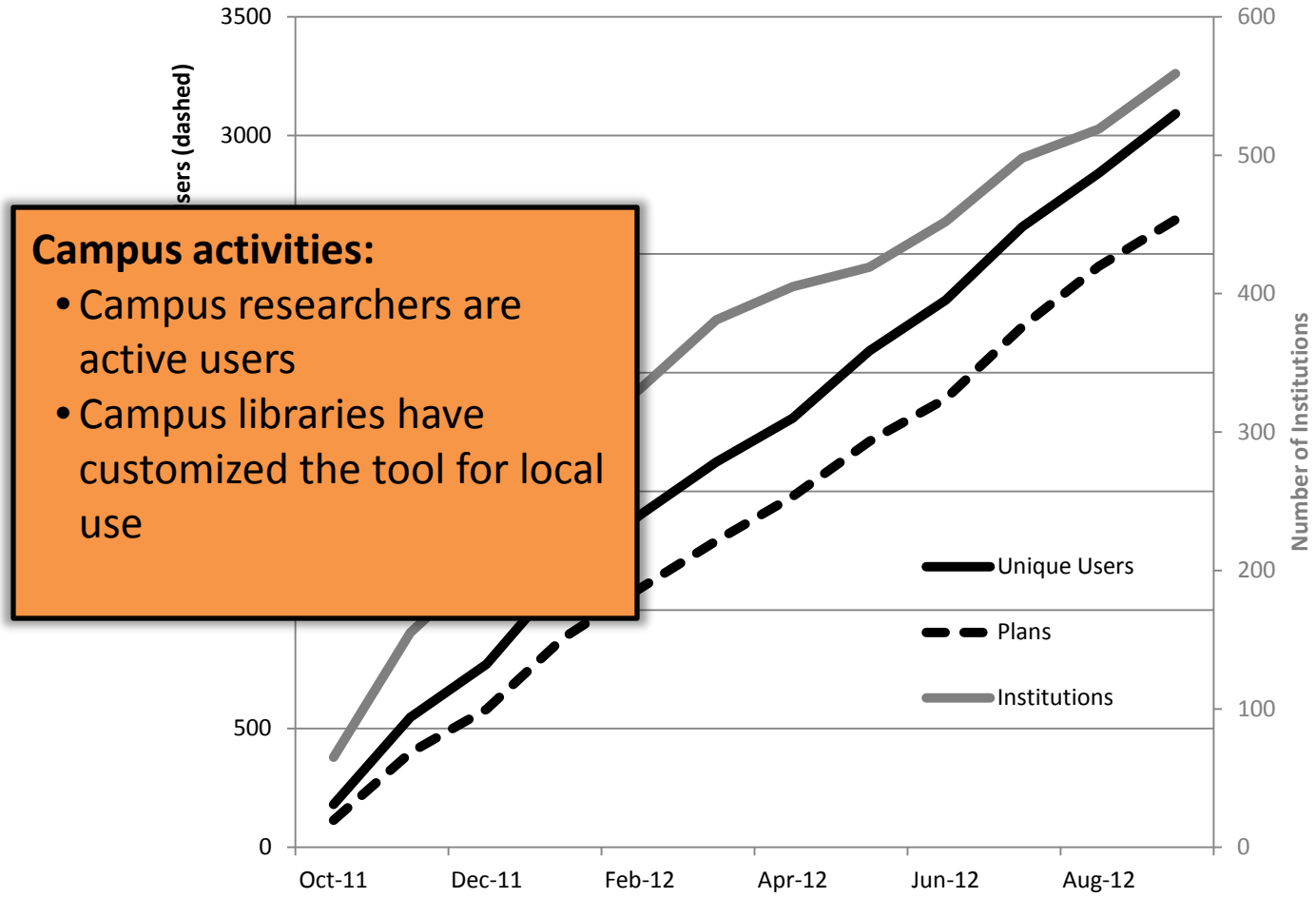
Primary Functions

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



UC3
UC Curation Center

DMTool usage



Campus activities:

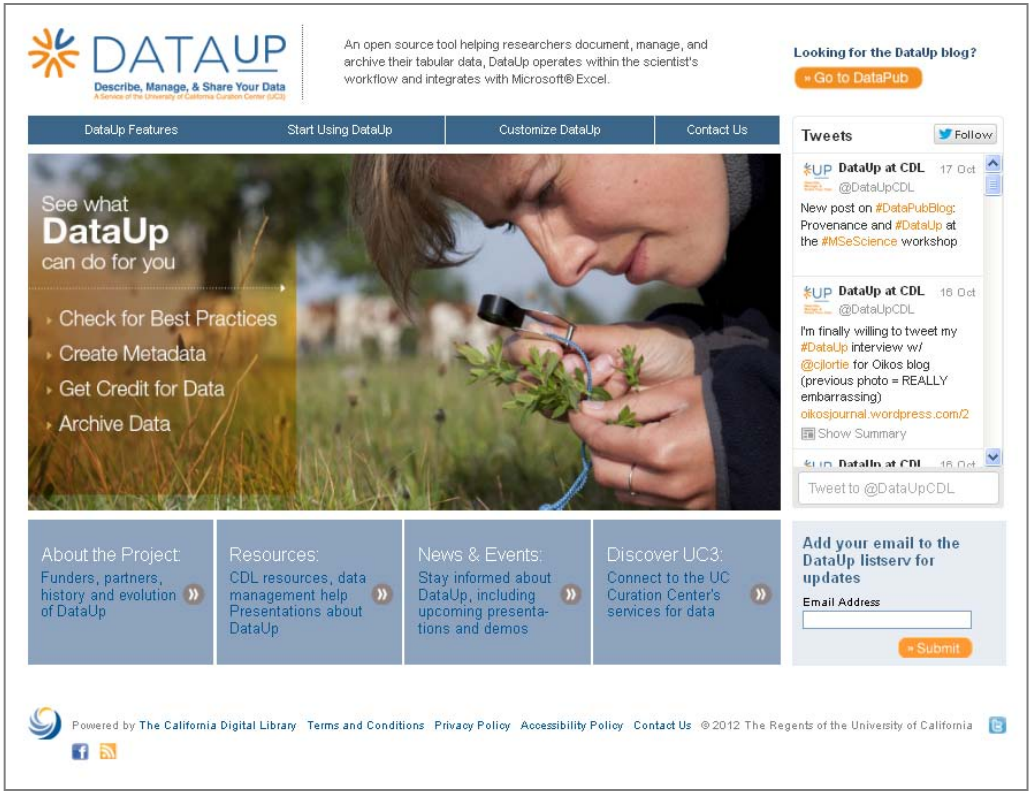
- Campus researchers are active users
- Campus libraries have customized the tool for local use

UC3 DATAUP
UC Curation Center

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



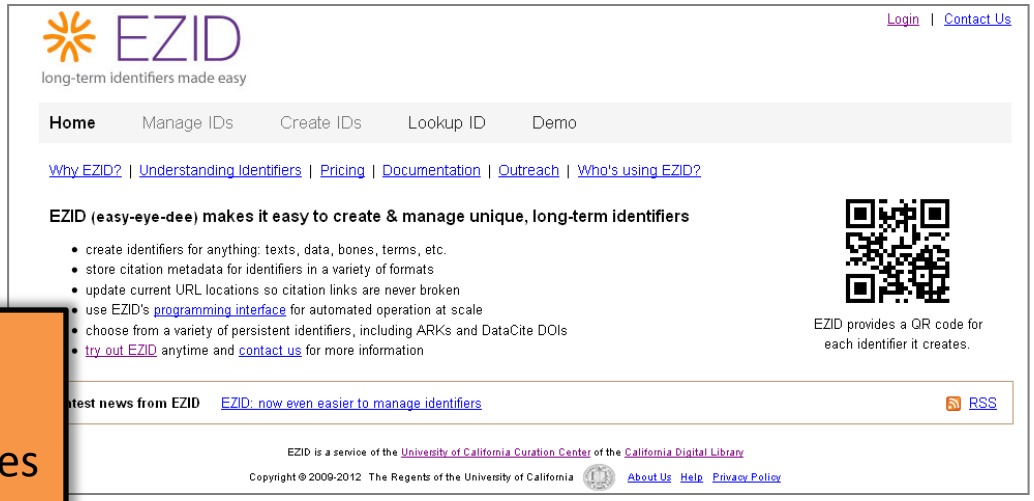
Microsoft Research

GORDON AND BETTY MOORE FOUNDATION

UC3 EZID
UC Curation Center

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 lite
- Ex
for
(W



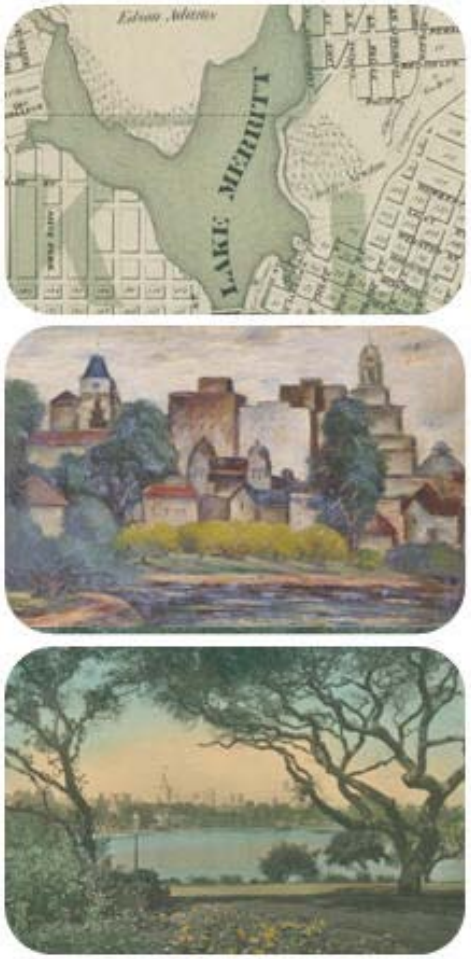
Campus library activities:

- 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

UC3 Merritt
UC Curation Center

Merritt Repository



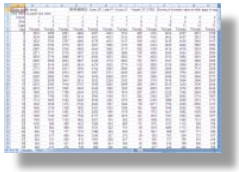
- Curation repository open to the UC community and beyond
- Discipline / content agnostic
- New kind of cost model: *pay as you go* or *pay once and store for X*

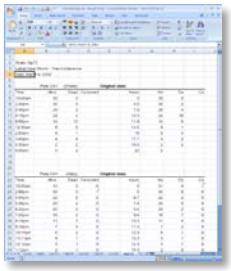
Primary Functions
1. Deposit
2. Manage (metadata, versions, etc)
3. Access (expose)
4. Share (with other researchers)
5. Preserve

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

ResearchHub
UNIVERSITY OF CALIFORNIA BERKELEY

islandora

eScholarship
University of California

omeka

Open Context

- Integration with distributed data grids

DataONE

DataShare

Open Data for the global scientific community

Find data

e.g. HIV, radiology Search

Share data

Upload research data

Browse by:

All Records

Lab/Department

Researcher

Learn more:

About DataShare

Why share research data?

FAQ

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.

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

This resource was made possible through a partnership between the California Digital Library, UCSF Clinical & Translational Science Institute, and UCSF Library.

University of California CDL California Digital Library

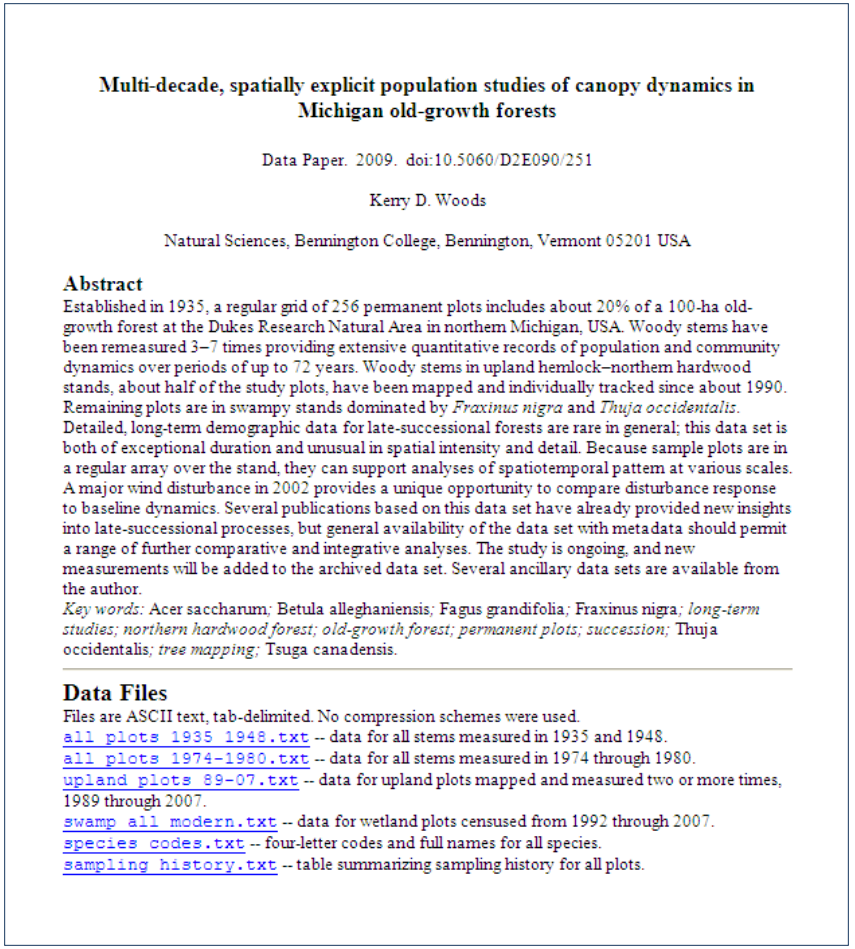
UCSF LIBRARY and Center for Knowledge Management

CTSI at UCSF

UC3
UC Curation Center

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 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.

GORDON AND BETTY MOORE FOUNDATION

UC3
UC Curation Center

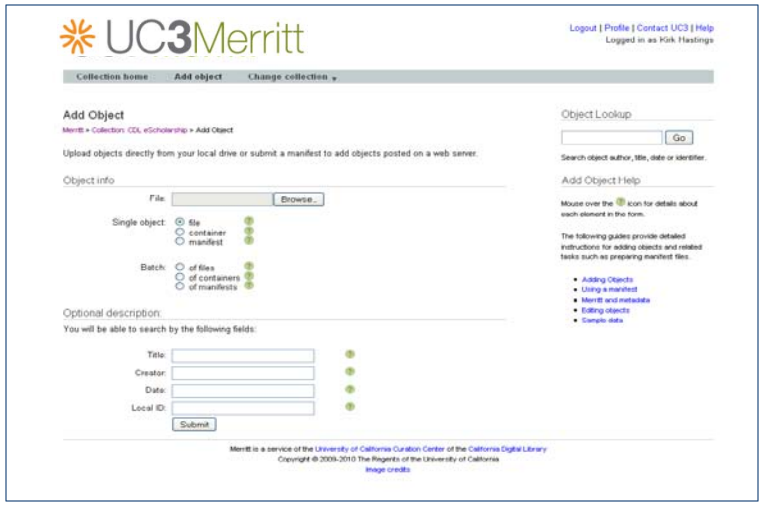
Data Publishing at the CDL

UC Curation Center

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

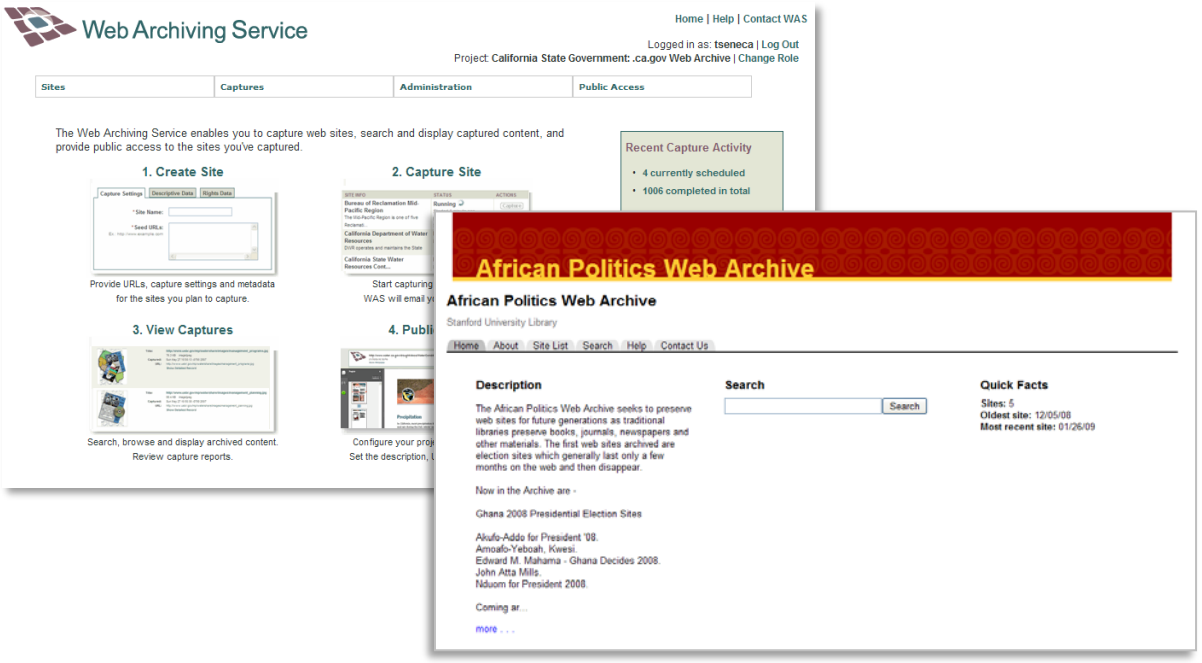
Publishing Services Program

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



UC3 WAS

University of California Curation Center Web Archiving Service



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

UC3
UC Curation Center

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

UC3
UC Curation Center

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/

UC3
UC Curation Center

For more information

UC Curation Center

http://www.cdlib.org/uc3

uc3@ucop.edu

Stephen Abrams
Patricia Cruse
Scott Fisher
Erik Hetzner
Greg Janée
John Kunze
Margaret Low
David Loy
Mark Reyes
Abhishek Salve
Joan Starr
Rosalie Lack
Carly Strasser
Marisa Strong
Adrian Turner
Perry Willett