

Project to Production: The Web Archiving Service

DLF Fall Forum, 2010 Tracv Seneca

> Web Ar	chiving Service	Projec	Logge t: California State Government: .ca.gov W	d in as: tseneca Log Out feb Archive Change Role		
ites	Captures	Administration	Public Access			
provide public acces	Service enables you to capture v s to the sites you've captured.	veb sites, search and display captu		apture Activity		
"Ste Name of Steel URLS, C	lu 🖟	State of Reciseration 166 and Parket Region 2 Parket Region 2 Transcription 166 and Parket Region 2 Transcription 166 and Parket Region 2 Transcription 166 and Parket Region 16	African Politic African Politics Web Arch Stanford University Library Home About Ste List Search H			
Search, browse	Captures Captur	4. Public Acces	Description The African Politics With Archive see whe sheets of future generations as tra libraries preserve books, journals, ner other materials. The first we bear an election after which generally last on incrition on the web and then dispose. Now in the Archive are - Ghana 2008 Presidential Election Sit Arusia, Addo for President '08. Amosto-Yebbah, Kwesi. Edward M. Mahmana - Ghana Decides John Jan Mille The State of President 2008. Coming ar	ddfional syspapers and chaed are yes few ways a few ar.	[Search]	Quick Facts Sites: 5 Oldest site: 12/05/08 Most recent site: 01/25/09





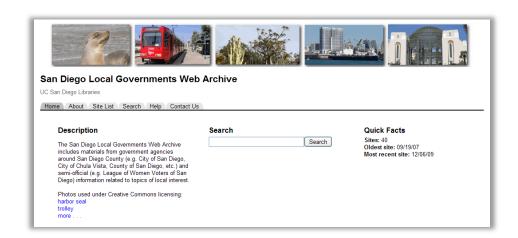


Web-at-Risk: 2005-2009

Today:

- Whirlwind tour
- Project to production
- Current / future work
- The DLF community

Build unique archivesfor local research communities



Geographically focused archives support local research

- Los Angeles
- Monterey Bay
- Orange County
- San Diego
- Santa Barbara

Topical archives support special research collections

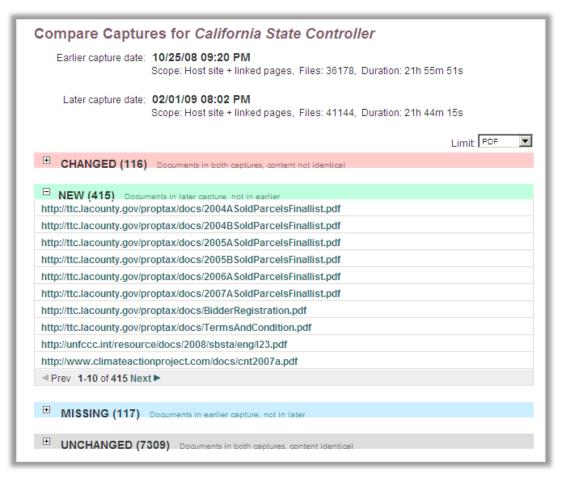
- Guantanamo Bay / Tamiment Library
- California Water Districts / Water Resources Center Archives



Archive Ephemeral Content



Analyze Site Change



Are there new documents on this site?

Are there documents in my archive that have been removed from the live site?

Easy to Use: Simple Workflow



Home | Help | Contact WAS

Logged in as: tseneca | Log Out Project: California State Government: .ca.gov Web Archive | Change Role

Sites Captures Administration Public Access

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.

3. View Captures



Search, browse and display archived content. Review capture reports.

2. Capture Site



Start capturing the site content.

WAS will email you when complete.

4. Public Access



Configure your project for public access. Set the description, URL or banner images.

Recent Capture Activity

- · 4 completed in past week
- · 670 completed in total

Capture Publication

This project is viewable by the public.

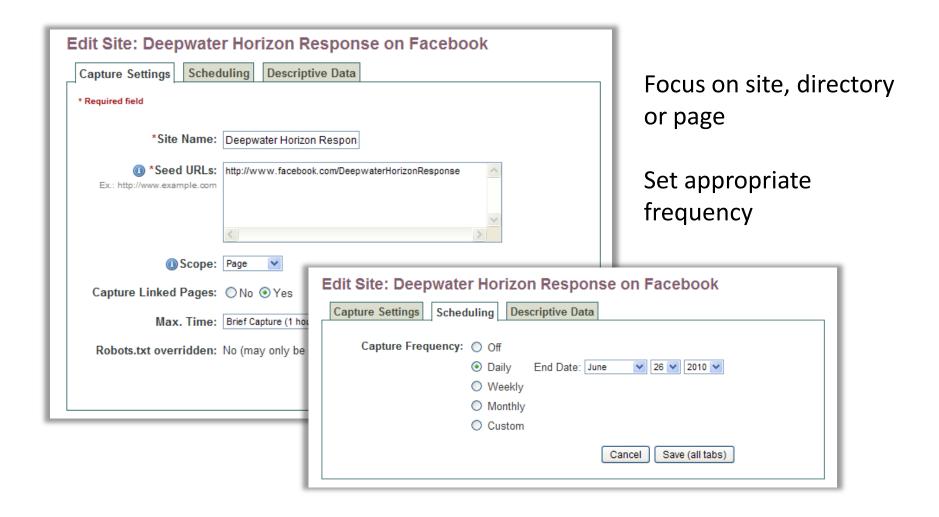
The public can view:

- 298 sites
- · 302 captures

Detailed Guides

- · Getting Started
- WAS User Guide

Tailor Capture Settings



http://was.cdlib.org



build an archive log in

Information for:

Partners

WAS News

Contact WAS

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.

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.

Learn more



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.

Learn more



Featured Archive

Santa Cruz Mountains Wildfire Web Archive



Created by: University Library, UC Santa Cruz

Sites: 35 Oldest site: 05/23/08 Most recent site: 05/29/08

Description: The Santa Cruz Mountain Wildfire web archive collection contains images, newspaper articles, and personal blogs related to the Santa Cruz Summit Fire of May 2008

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

2009 H1N1 Influenza A (Swine Flu) Outbreak

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

University of California Curation Center, California Digital Library

Archive Home Page



California State Government: .ca.gov Web Archive

UC Government Information Librarians

Home About Site List Search Help Contact Us

Description

The ca.gov domain is the 3rd largest domain in the U.S. government web. A great deal of that content is published by the State of California, and represents a rich source of material for researchers, students, the general public and state agencies themselves. This archive preserves access to hundreds of state agency sites that are captured on a regular basis.

State agencies utilize their websites to publish everything from press releases, agendas, minutes, events, reports and statistics. This materia...

more . . .

Search

Search

Quick Facts

Sites: 299

Oldest site: 10/24/08 Most recent site: 01/08/10

Automatic List of Sites











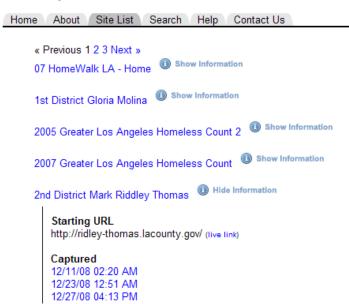






Los Angeles Local Government Web Archive

UCLA Library



lookup by site name Go Clear

Click the "information" icon to see further information about the site, including the dates on which each site was captured.

The site lookup box will find words anywhere in the site name.

Search Across All Sites or Target Just One











San Diego Local Governments Web Archive

UC San Diego Libraries

Searc	:h			
Ţ	ype words or a URL:	_	Filter by:	
b	order		All sites	Search
« Previo	ous 1 2 3 4 5 6 7 8 9 10 Next »		All file types	

Your search terms will be found anywhere in the full text of the web pages and documents in this archive. You can search for key words or for particular URLs.

You can narrow your search to find the words in particular web sites, or to find only particular file types, such as PDF files.

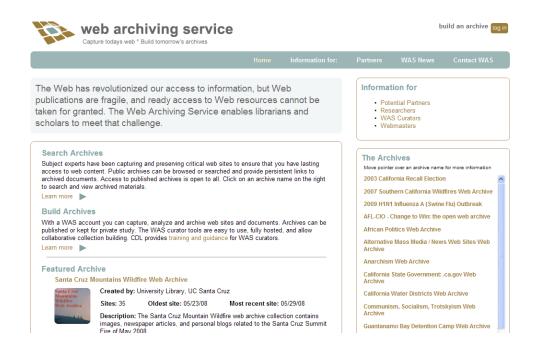
You are not required to type a key word in your search; you can select a site to review all of the documents from that site.

Border Affairs - SignOnSanDiego.com
Captured: Sun Dec 06 17:31:50 UTC 2009
www.signonsandiego.com/news/border-affairs/ 66.8 KB text/html

Mexico-U.S. Border: AFSC
Captured: Wed Apr 30 20:05:10 UTC 2008
www.afsc.org/mexico-us-border/default.htm 17.3 KB text/html
... foundations of effective cross-border organizing.
Committee of Workers ...

The Border
Captured: Sat May 17 12:04:20 UTC 2008
www.pbs.org/kpbs/theborder/history/hdrhistory.html 527 Bytes text/html

WAS Snapshot: October 2010



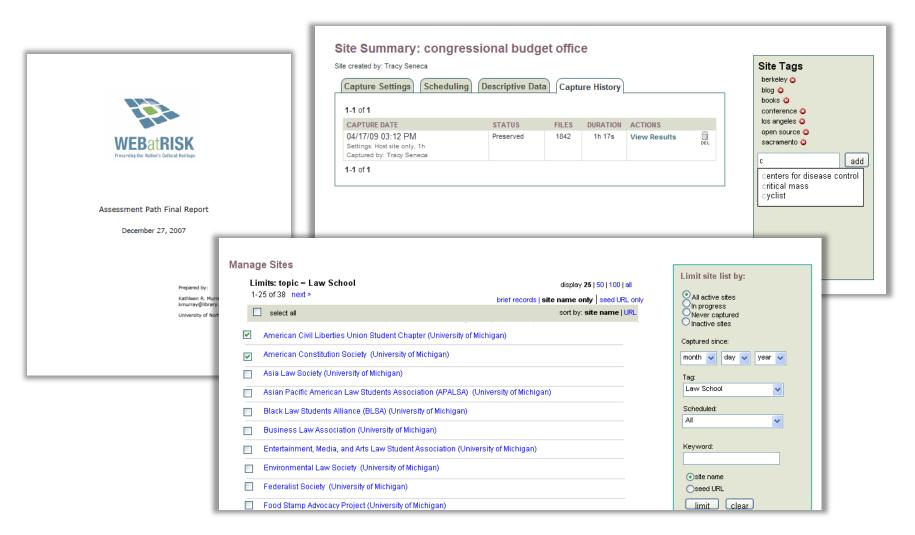
Stats: Since January 200717 organizations4,509 sites captured33,619 captures run21.7 terabytes29 archives published

Project to Production

- Scale
 - Infrastructure: ensure continuity of service through server upgrades
- Policy
 - Setting up service agreements

Engage the users

(Needs assessment doesn't stop)



University of California Curation Center, California Digital Library

"Production" is not just for us!

- How does WAS fit into existing workflows?
- Is there a 'web archivist'? Student workers?

Friday November 5th 10-11:30 am

Curators from UC San Diego, NYU, University of Michigan and UC Berkeley share their approaches to web archiving.

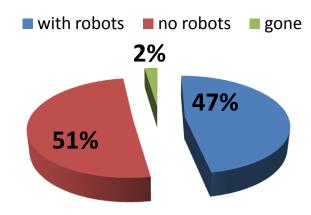
http://was.cdlib.org WAS News



Archiving the State of California

Learning from the archives

California State Agency Sites



307 sites 1363 captures 2.7 TB Began October 2008



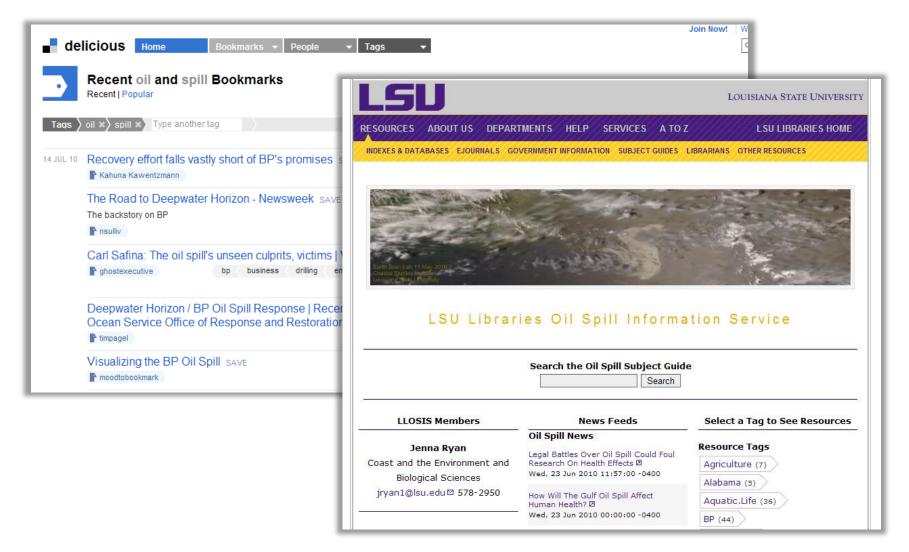
Archiving the Gulf oil spill

Improving support for collaboration

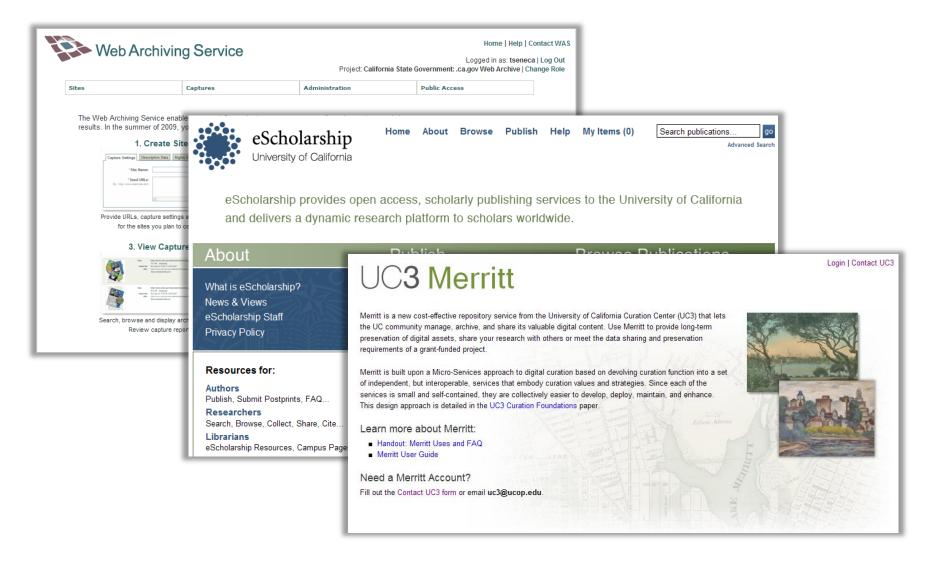


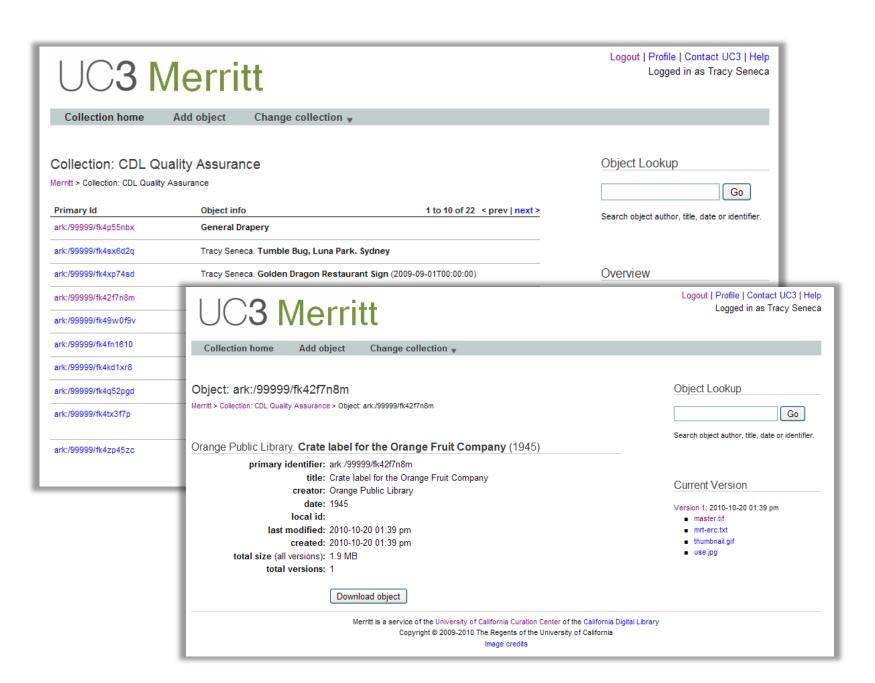
527 sites 9288 + captures 1.6 TB Began May 5

LSU tags relevant sites in Delicious CDL imports Delicious JSON feed into WAS



Interoperability of CDL Services





DLF Community

Tools

- WAS
 - UC libraries, NYU, Stanford
- Archive-It
 - Columbia, Rice, Indiana...
- Other
 - Library of Congress, North
 Texas, Los Alamos

Common issues/developments

- Memento
- Policy
 - Can / should Facebook block government pages from capture?
- Collection
- Collaboration
- Imagination

In-depth demos: Nov 8th and 16th Workflows discussion: Nov 5th

- http://www.facebook.com/webarchiving
- http://was.cdlib.org > WAS News

