Web Archiving Service
Capture Today's Web • Build Tomorrow's Archives

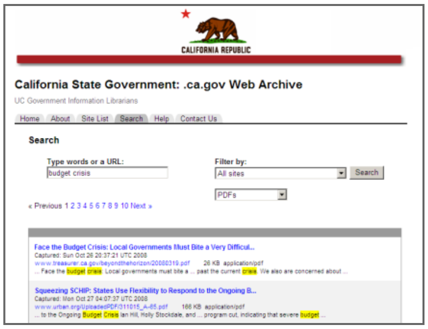
http://was.cdlib.org

Archiving the Web

The web has revolutionized our access to information. Documents and publications that were once difficult to fin are now readily available to anyone. Government agencies, non-profit organizations and other critical sources now have an inexpensive means for distributing information to the public. When important social and political events take place, we can see the public reaction unfold via blogs and personal web sites, and have an unprecedented view into popular culture and the debates that shape public policy. All of these materials will serve as valuable resource for researchers for years to come.

But ready access to these publications cannot be taken for granted. When sites are redesigned, when new administrations take office, when policies or organizations change, we witness the wholesale disappearance of information.

The California Digital Library offers the **Web Archiving Service (WAS)** to help you meet these challenges.



California State Government: .ca.gov Web Archive

UC Government Information Librarians

Search

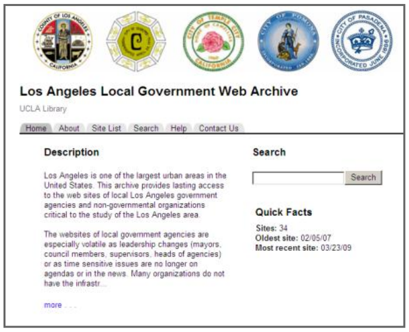
Type words or a URL: budget crisis

Filter by: All sites

« Previous 1 2 3 4 5 6 7 8 9 10 Next »

Face the Budget Crisis: Local Governments Must Bite a Very Difficul...
Captured: Sun Oct 26 20:37:21 UTC 2008
www.treasurer.ca.gov/beyondthehorizon/20080319.pdf 26 KB application/pdf
... Face the budget crisis: Local governments must bite a ... past the current crisis. We also are concerned about ...

Squeezing SCHIP: States Use Flexibility to Respond to the Ongoing B...
Captured: Mon Oct 27 04:07:37 UTC 2008
www.urban.org/UploadedPDF/311015_A-65.pdf 166 KB application/pdf
... to the Ongoing Budget Crisis Ian Hill, Holly Stockdale, and ... program cut, indicating that severe budget ...



Los Angeles Local Government Web Archive

UCLA Library

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.

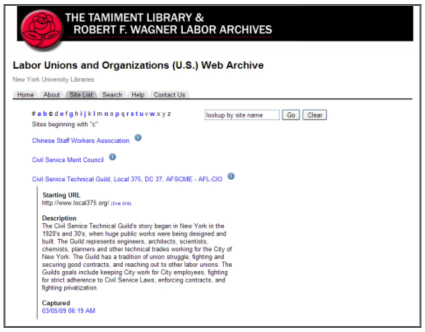
The websites of local government agencies are especially volatile as leadership changes (mayors, council members, supervisors, heads of agencies) or as time sensitive issues are no longer on agendas or in the news. Many organizations do not have the infrastr...

more . . .

Search

Quick Facts

Sites: 34
Oldest site: 02/05/07
Most recent site: 03/23/09



Labor Unions and Organizations (U.S.) Web Archive

New York University Libraries

Sites beginning with "c"

Chinese Staff Workers Association

Civil Service Merit Council

Civil Service Technical Guild, Local 375, DC 37, AFSCME - AFL-CIO

Starting URL
http://www.local375.org/ (live link)

Description
The Civil Service Technical Guild's story began in New York in the 1920's and 30's, when huge public works were being designed and built. The Guild represents engineers, architects, scientists, chemists, planners and other technical trades working for the City of New York. The Guild has a tradition of union struggle, fighting and securing good contracts, and reaching out to other labor unions. The Guilds goals include keeping City work for City employees, fighting for strict adherence to Civil Service Laws, enforcing contracts, and fighting privatization.

Captured
03/05/09 06:19 AM

What you can do with WAS

Build unique archives for local research communities

You can easily create resources of immediate and lasting value to your local research community by archiving local government agencies and public policy organizations. These resources are not likely to be available to historians by any other means. Researchers can search all of the sites you archive together or individually, and you can customize the archive with your own branding and imagery.

Archive web content for study and analysis

Researchers may need to study unique data sets, and to study groups of sites that no one else has archived in one place. WAS archives can be created and maintained for private study and analysis as well as public access. WAS provides tools for analyzing site change over time and allows keyword searching for archived sites, and publishing an archive is optional. WAS can also serve researchers by providing lasting access to ephemeral web sites resulting from grant activity.

Preserve your own organization's web presence

What if you could easily go back and compare your services and publications of five years ago to those same services today? Whether you are a small non-profit organization or a major University, even short-term management of your web content can be challenging, and keeping web content current is often a high priority. Very large, complex organizations may have equally complex web networks consisting of hundreds of related sites. Whether simple or complex, WAS provides tools to archive your web presence on a periodic, scheduled basis to preserve a record of your organization.

Web Archiving Service Features

Easy to use You don't need specialized knowledge of web archiving to use the service; all you need is the subject expertise that you already have. The WAS interface is intuitive and user-friendly. If you do need help, the California Digital Library provides customer support, as-needed guides and in-depth training sessions.

Fully hosted service You don't need a storage infrastructure or dedicated IT staff to run web crawlers. The California Digital Library hosts both the service and storage with complete data center support.

Flexible, collaborative accounts You decide which users can contribute content to the archives you create. If you need to collaborate across institutions or forge collaborations between librarians and faculty, WAS will support your needs.

Tools for analysis The Web Archiving Service offers unique tools to evaluate your web captures, including the ability to analyze how sites change over time.

Curatorial control Not all web sites are the same; web archives can be very different depending on the nature of the sites you target. With the Web Archiving Service, you can tailor your capture settings and frequency to individual sites as needed.

Branding You can easily tailor your archives to integrate with your own web site. You can also create an interface that highlights the unique resources within your archive.

How WAS accounts work

You will have administrative tools to manage your users and archives. You can create as many separate archives as you wish.

University of California The California Digital Library provides digital library services to the University of California. UC departments and organizations are charged only for the storage used.

Non-UC Institutions Non UC Institutional accounts are charged a yearly service fee and begin with 1 terabyte of allotted storage. Additional storage can be purchased as needed.

Consortia Discounted service fees are available for consortia of three or more institutions.

Contact washelp@ucop.edu to inquire about an account.

> *"[Middle East researchers] said that they wish such services had been available years ago as many web sites of important political groups have long disappeared, having been shut down or abandoned. An example is, of course, all of the web sites from Iraq during the Saddam Hussein era. Valuable research information on the inner workings of the state and its various departments are lost forever. This information would be invaluable to determine even the current situation in this still unstable country"*
> *- John Eilts, Stanford University*

Archiving California’s Web

The University of California Provides Lasting Access to California State Websites

http://webarchives.cdlib.org/calgov

The web has revolutionized our access to information. Open government initiatives coupled with the ease of Web publishing provide an unprecedented view into the research and debates that shape public policy. This is particularly true in California, which is not only the world’s 8th largest economy, but is also the U.S. government’s 3rd largest Web domain. Unfortunately, ready access to California’s rich Web cannot be taken for granted. The Web is ever-changing and maintaining an archival record of websites is a challenge. In 2008, the California Digital Library (CDL) began capturing State of California websites using CDL’s Web Archiving Service. The California State Government Web Archive includes over 300 state agency sites that are captured twice a year with more frequent captures of critical sites identified by the University of California Government Information Librarians’ group. The archive currently holds 2 terabytes of data (13,387,140 files) and can be searched by keyword or URL and browsed by site name.

Government Transparency and Ready Public Access

The California State Government Web Archive preserves the legacy of administrations as they change and provides lasting access to agency sites when state government offices are restructured. The archive also holds ephemera of interest to public policy researchers such as public meeting minutes and agendas which are often removed from Web sites to keep them up to date.

The public can access and study archived sites in ways that aren’t possible on the live Web. An archive can show which agency sites are the largest, which offer the most multimedia, and which change most frequently. A recent study of State of California Website Trends 2008-2010 provides an example of what can be learned when a government Web presence is used as a data set for study.
http://www.cdlib.org/services/uc3/docs/cagovsites.pdf

California Digital Library

The CDL was founded by the University of California in 1997 to take advantage of technologies that were transforming the way digital information was being published and accessed. Since then, together with the UC libraries and other partners, we have assembled one of the world’s largest digital research libraries and changed the ways that faculty, students, and researchers discover and access information.

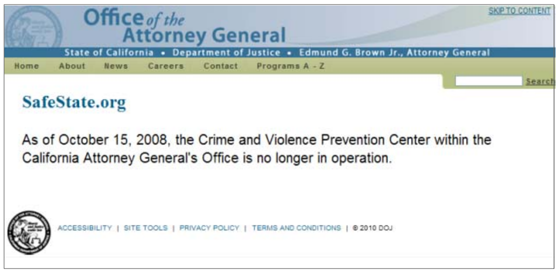
http://www.cdlib.org

The Archive Provides Value to State Agencies

In an environment of fiscal challenges, many state agencies may not have the resources to maintain an archival record of their digital publishing activity. As the composition of State government changes and as sites are redesigned and updated, the California State Government Web Archive can provide a lasting record of State publications both to the general public and to the agencies themselves.

Case: California Crime and Violence Prevention Center

Live Web, May 2010

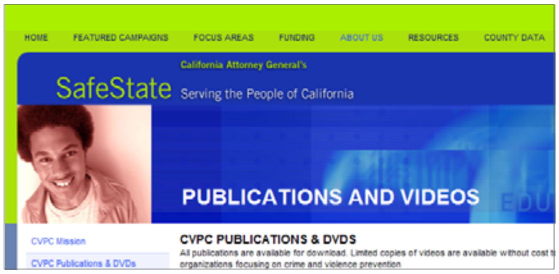

SKIP TO CONTENT
Office of the Attorney General
State of California • Department of Justice • Edmund G. Brown Jr., Attorney General
Home
About
News
Careers
Contact
Programs A - Z
Search

SafeState.org

As of October 15, 2008, the Crime and Violence Prevention Center within the California Attorney General's Office is no longer in operation.

ACCESSIBILITY | SITE TOOLS | PRIVACY POLICY | TERMS AND CONDITIONS | © 2010 DOJ

The Archive



CVPC PUBLICATIONS & DVDS

All publications are available for download. Limited copies of videos are available without cost t organizations focusing on crime and violence prevention.

The archive provides access to thousands of HTML pages, over 500 PDF publications, and to video and audio files from the now retired Crime and Violence Prevention Center site.

Challenges to Saving the State's Web

Web Archiving can be challenging, particularly as sites become more multimedia rich and interactive, but the greatest barriers to archiving Web content are not technical, but policy issues.

One aspect of Web site management that profoundly affects Web archiving is a server standard called *robots.txt files*. These are files on a Web server that provide instructions for Web crawlers. Site owners have a good deal of control over what they can specify in robots.txt files. They can decide to limit access to particular files, they can prohibit entire directories from capture, or they can entirely prevent crawlers from capturing the site. These rules primarily ensure server performance, but they also severely limit the ability to archive a site.

Web Archiving Service

The California Digital Library provides the Web Archiving Service (WAS) to enable archivists and researchers to capture websites and build publicly accessible archives. WAS is easy to use, offers flexible settings to help you capture sites effectively and provides tools to search and analyze archived web content.

http://was.cdlib.org

The California Digital Library's crawlers comply with robots.txt files. As of September2009, 47% of California agency sites had robots.txt files. Of those, 4% entirely prevent the site from capture, while more than half cause the archival copy of the site to render poorly. In these cases, images are often missing, the site's format does not display correctly and major portions of content may be missing. This includes sites that are critical to the study of the State of California such as the Senate Office of Research, the California State Controller, and the Secretary of State.

Potential Collaboration: A Better Archive (and Better Websites!)

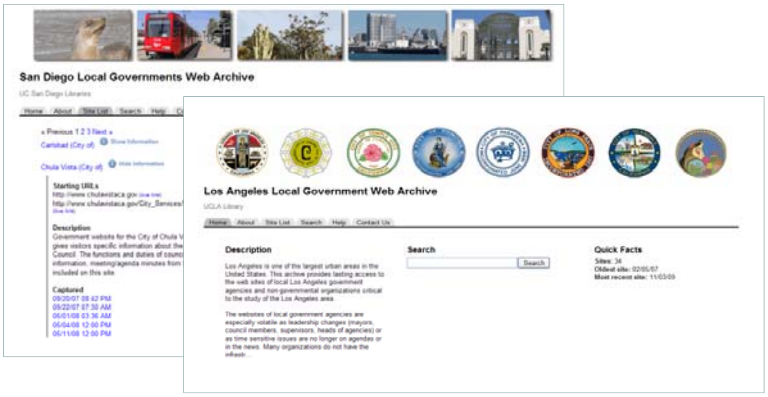
To help create a more comprehensive and historically accurate archive, state agencies can expressly permit the California Digital Library to archive their sites by adding the following two lines to the top of their robots.txt files:

```
User-Agent: cdlwas_bot
Disallow:
```

There are even greater opportunities to benefit both the public and the state agencies. Web servers can be configured to provide error messages that are linked to the archive. If a user encounters a 404 File Not Found page, the error message itself can provide a link to the archived copy of that page. With this kind of collaboration between content owners and archives, the public would not even have to know about an archive to benefit from it.

Other California Web Archives at the California Digital Library

In addition to state agency sites, curators at the UC campuses are building vast collections of additional California information. Some, such as the California Water Districts Archive, are focused on a topic, while most are regional archives of county and city websites. These actually surpass the California State Government archive in size and scope. CDL supports 17 other California Web archives totaling over 1,200 websites and amounting to 5 terabytes of data. Eleven of these are available to the public (listed below). In addition, CDL had begun archiving websites before the Web Archiving Service was available.. This includes government agency content from 2005 – 2008 and the 2003 California Governor Recall Election Archive, which is available to the public.



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 especially volatile as leadership changes (mayors, council members, supervisors, heads of agencies) or as time sensitive issues are no longer on agendas or in the news. Many organizations do not have the infrastr...

- 2003 California Recall Election
- 2007 Southern California Wildfires
- California State Government
- California Water Districts
- Los Angeles Local Government
- Monterey Bay Area Local Government
- Orange County Government
- San Diego Local Governments
- Santa Barbara Government
- Santa Cruz Mountains Wildfire
- Ventura Government

Questions? Contact Patricia.Cruse@UCOP.EDU; Tracy.Seneca@UCOP.EDU