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.

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.

May, 2011

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. The cost for storage is $1040 per terabyte, an average of $1.46 for a comprehensive website capture.

Non-UC Institutions Non UC Institutional accounts are charged a $10,000 yearly service fee and begin with 1 terabyte of allotted storage. Additional storage can be purchased as needed at $1,040 per terabyte.

Consortia Consortia consisting of three or more institutions qualify for a $1,000 discount on the yearly service fee. Each institution receives a separate account.

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