



# California Digital Library's Web-based Government Information Project FAQ

## ***What is the Web-based Government Information Project?***

The California Digital Library is conducting a cost-benefit review of technologies appropriate for the capture, curation, and persistent management of web-based documents of US state and federal governments. Project activities include:

- Analyze the scope of the domain of web-based government information
- Identify promising technologies and projects related to capture, management and preservation of this material
- Capture a subset of web documents from government sites to serve as a testbed for analyzing the technical requirements for building collections of web-based government documents

## ***What is a “Memory Organization”?***

A *memory organization* is any institution whose mission is to preserve the cultural legacy of societies for future generation. Libraries, museum and archives are examples of memory organization that have traditionally focused upon the preservation of content created by among others, writers, artists, scientists, scholars, government agencies and commercial organizations.

## ***What is “Curation”?***

*Curation* is an activity that is core to the mission of memory organizations. The OED defines the verb curate thusly: “To act as curator of (a museum, exhibits, etc.); to look after and preserve”. Curatorship is often used within the library community to describe professional staff involved in the long-term guardianship and preservation of rare books and manuscripts.

## ***What is “Web-based Government Information”?***

The focus of this project is content that is produced by government agencies or government officials and published exclusively on the web or made accessible via web-accessible interfaces. Although the project is initially on preserving static content, it is also interested in the preservation challenges posed by dynamic databases.

## ***Isn't this work the responsibility of government agencies?***

It would be wonderful if all government agencies had the resources and conscience to retain and preserve their entire intellectual output in perpetuity. In reality, few agencies possess the means or inclination to do this.

***Isn't this work the responsibility of government archives and libraries?***

Yes. The mission of government archives and libraries generally include the preservation of cultural records, documents and publications. However, these organizations face the same resource constraints as all government agencies. Memory organizations outside of government have traditionally played a critical and often leading role in the preservation of cultural content in the era of print.

***How does this project address the issue of authenticity of web-based government information?***

Emerging technologies and metadata strategies can be adopted to ensure then authenticity of web-based government information irrespective of the server on which it resides. In addition, non-governmental memory organizations have historically performed a role as trusted repository for government information, by virtue of its independence of political imperatives and management. The volatility of digital content suggests that memory organizations outside of government should assume a leading role to ensure that government records, documents and publications are not manipulated or withdrawn to achieve politically-motivated objectives.

***Who is involved in this project?***

This project is a highly collaborative project supported by funds from the Andrew W. Mellon Foundation. The lead organization is the California Digital Library, where analysis of the government information domain and review of key projects are being conducted. The development of a testbed collection of web-based government documents involves collaboration with the faculty and staff at both the Stanford Computer Science Department and the San Diego Supercomputer Center.

***What agency websites have been crawled in order to create the project testbed?***

Six sites have been crawled including: United States Department of State; United States Geological Survey; United States Senate; California State Water Resources Control Board; California Energy Commission; and the California Legislative Analyst's Office.

***What have you accomplished thus far?***

Surveys of several leaders in the field as well as managers of over 25 projects are being conducted. A report summarizing the results of this work will be available by early March. Two crawls have been conducted and the content is being ingested into a repository. A report on web metrics of web-based government information based upon a literature review and analysis of the crawl will be available in early March. A technical review of available technologies to support capture, curation and preservation of the content is underway.

***What are you planning to do next?***

Project staff will be presenting their reports to meetings of various stakeholder groups in the coming months, including sets of librarians who serve as intermediaries in identifying and retrieving government content, as well as representatives from key user groups including faculty, researchers, and graduate students. The goal of these meetings will be the development of a set of functional specifications for the technical infrastructure necessary to capture this genre of content as well as specification for providing a service layer surrounding this infrastructure.