

Serving Up California Government Data: An Integrated Approach

Patricia Cruse(*), Marsha Fanshier(*), Fredric Gey(+),

* California Digital Library
Office of the President
University of California
1111 Franklin Street, 7th Floor
Oakland, CA 94607-5200

+ UC DATA Survey Research Center
University of California
2538 Channing Way #5100
Berkeley, CA 94720-5100

patricia.cruse@ucop.edu, mfanshier@ucsd.edu, gey@ucdata.edu,
<http://countingcalifornia.cdlib.org>,

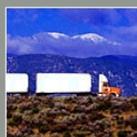
Counting California Is:

A new initiative committed to enhancing California citizens' access to the growing range of social science and economic data produced by government agencies. In a departure from more static formats, Counting California's single interface enables users access to actual raw data compiled by federal, state, and local agencies, and also allows users to collate and integrate data by topic, geography, title, and provider. Counting California also addresses the serious preservation dilemma posed by ever-changing technology and data formats, insuring easy and continuous access to historical and current information. Counting California is available at: <http://countingcalifornia.cdlib.org>.



Counting California

Quick Facts: [California](#) | [Counties](#)



[Home](#) [About Us](#) [Feedback](#) [News](#) [Help](#)

Information, facts, and data about the Golden State

Explore Information by: [Topic](#) | [Geography](#) | [Title](#) | [Provider](#)

Monthly Poll

What is your usual method of transportation to work?

- Drive alone
- Carpool
- Bus
- Subway
- Bicycle
- Other

[View Results](#) [Get Data](#)

Highlights

- **by Topic:** Check out the incredible array of topics for which we have the numbers — from the value of the state's anchovy fishery to...well...take a look!
- **by Geography:** Focus on an issue for a particular area — by county or city or a combination
- **by Title:** Data, we've got data! This tells you which piles of data (called datasets) we have
- **by Provider:** Here we list the data by the institutions providing it

The Change in Access to Government Information -- Rising to the Challenge

Government data serve as varied a clientele as there are Californians, ranging from businesses and private citizens to some of the most prominent educational and research institutions in the world. With a population and economy larger than those of many European countries, the state of California's need for continuous and uniform access to government-produced data has never been more urgent.

Ironically, though, the digital technologies that in recent years have revolutionized data distribution have also created unintentional problems. What was once a stable system of print materials has given way to a diffuse, constantly changing array of electronic media, each using different formats and access methods. The current climate leaves many would-be users frustrated or bewildered; each new upgrade of software and web browsers only exacerbates the problem.

Preservation and consolidation of historical, or time-series, data are similarly at risk. Government agency web sites often mount new information, but may follow no systematic plan to preserve older, historical data as each update supersedes the last. The lack of cumulative time-series data can effectively cripple any attempt to discern long-term trends and changes.

Goals of Counting California

The need for uninterrupted data access and the preservation of historical data are the two biggest problems facing government data distribution today. Counting California's goals respond directly to these challenges:

- To provide flexible, user-friendly access that meets the diverse needs of the California citizenry;
- To insure uniform, continuous access to both current and historical government data; and
- To foster the ability to share data and work collaboratively between government agencies and other members of the data community.

Technology, Innovations, and Core Components

One of Counting California's primary aims is to encourage collaboration, and the ability to share our work with data producers and other members of the data community. Accordingly, we have adopted several measures to promote universal standardization:

Data Documentation Initiative (DDI)

For the past several years the social science data community has been working on the Data Documentation Initiative (DDI) to produce a metadata standard for social science data resources. The DDI "is an effort to establish an international criterion and methodology for the content, presentation, transportation, and preservation of 'metadata' about the datasets in the social and behavioral sciences." The DDI committee has produced what is known as a Document Type Definition (DTD) as the foundation of its data discovery and display system.

Counting California has made extensive use of the DDI and its extension to include "matrix variables" to act as an intermediary between storage and display of over 250 statistical tables about California cities and counties. Our experience is that centralized metadata facilitates flexible thinking about data storage, data discovery and data display. This paper will describe our experiences and results, as well as our plans for future enhancements. This project has taken a major step in moving the DDI beyond its archival role to play an active part in control of the data storage, discovery and display processes of statistical information systems.

XML and DDI

The DTD employs the eXtensible Markup Language (XML) as a platform for the DDI standard. XML, a dialect of the more general SGML markup language, is used for documents containing structured information. In recent years there has been explosive growth in the use of XML and web-based XML tools including databases, search engines, and editors.

In keeping with our goals, we intend to make our DDI DTD metadata available to the scholarly community. Links to actual XML will be available from the web site. We will encourage others to build upon our work in creating alternate applications that can also be shared.

SAS IntrNet

The backbone of Counting California's data delivery capacities is SAS IntrNet, which allows for the integration of SAS and the World Wide Web. Specifically, SAS allows for data extraction via the metadata database and provides the ability to format the data in tables, charts, maps, and graphs. In the future we hope to take full advantage of SAS IntrNet's capabilities as we add functionality to the system.

Open Communication

Throughout the development of Counting California, the project team has relied heavily on the expertise of the data community. DDI committee members Ann Gerken Green (Yale University), Wendy Treadwell (University of Minnesota), and Cavan Capps (Bureau of the Census) provided guidance in the early stages of the project by answering our many questions about the DTD. Mary Vardigan (Interuniversity Consortium of Political and Social Research) furnished reference support as we struggled to collect examples.

Counting California Partners

Counting California is a collaborative project funded by the California Digital Library and the Library of California. A portion of the funding for Phase I of Counting California comes via special arrangement with the Library of California. This interagency agreement stipulates that the Library of California provide funding, and the CDL implement the project. Some of the funding for Phase II comes through a grant from the Library Services and Technology Act (LSTA), administered by the California State Library.

Current strategic partners in data acquisition include the U.S. Bureau of the Census, and the California Department of Finance. We are currently negotiating similar arrangements with numerous other state agencies.