July 2014

OSU creates web site for government statistics

Carolyn Ottow
Oregon State University

OSU Creates Web Site for Government Statistics

by Carolyn Ottow
Oregon State University

In the early 1990s, as the federal government began publishing more and more information on CD-ROM, depository libraries faced many technical challenges. Figuring out how to mount the dozens of disks was the first hurdle. The 1990 Census alone was issued on more than 100 CDs. Clunky software, inconsistent user interfaces and the inability to network some CDs also presented problems. Despite difficulties, librarians at Oregon State University, who had already had success developing a network for commercial CDs, saw this as an opportunity for outreach. With data in electronic format, surely there was a way to provide remote access to libraries around the state that did not otherwise have easy access to these government databases. With this in mind, Charlene Grass, OSU's associate university librarian for technical services, wrote a grant proposal to network government CDs, and the Government Information Sharing Project was born.

The Project began in 1993 with funding from the U.S. Department of Education. The intention was to create a CD network that would be accessible via telnet and to work with software developers to create a more intuitive, consistent user interface to the many and varied CDs. As the Wide Web emerged as the new standard for Internet access, it became clear that using a Web site would provide the best access to the CDs. The Web allowed us to develop a consistent interface that is easily accessible through popular browsers like Netscape and Mosaic. Using standard features of the hypertext markup language (HTML) used to write Web pages, staff on the project created interactive forms and clickable maps that allow users to extract data tables from several different databases. The URL for the home page is govinfokcr.orst.edu.

The resources on the site include some of the most useful and popular government statistics in the areas of demographics, economics and education. The first database introduced on line was the 1990 Census data for Oregon and other Northwest states. This has recently been expanded to cover all 50 states. Other demographic databases are Population Estimates for counties from 1990-1992 and USA Counties, a handy reference source with social, economic and governmental statistics spanning several years.

The economic databases include the 1992 Census of Agriculture, 1992 Economic Census, U.S. Imports and Exports, and the Regional Economic Information System (REIS) from the Bureau of Economic Analysis, which has income and employment data as well as short narrative summaries of regional economies. The recently added Consolidated Federal Funds Report details federal expenditures in all states, counties and municipalities.

Finally, the School District Data Book provides a wealth of demographic, financial and administrative data for all U.S. school districts. Users can find such things as the student-to-teacher ratio in their school districts and the level of funding by federal, state, and local governments.

Although these data bases are easy to use, providing access to them is labor intensive. All of the CDs contain files in dBase III format. By adapting DButil software developed at Lawrence Berkeley Laboratories, programmers on the project wrote software to read and format the data files. CGI scripts were written for the Web pages’ interactive forms. When a user queries the site, the programs extract the data directly from the CDs, which reside in drives attached to the Web server. One exception is the 1990 Census data. Because of the large number of CDs, these data files were extracted, subsets, and stored on the server’s hard drive. We have found that getting the data directly from the CD-ROM is almost as fast as accessing the dBase files directly from the hard disk.

Project staff members have taken care to make the Web site easy to use. Keyword searching is available within each database, allowing users to pinpoint statistics on specific topics. Besides the numerical data, the complete documentation from each CD was coded in HTML and is available on the site by clicking on the “info” buttons. Users can find out how the data was compiled, look up definitions of terms, and read about sources and authority of the data. The URL for the home page is govinfokcr.orst.edu.

The resources on the site include some of the most useful and popular government statistics in the areas of demographics, economics and education. The first database introduced on line was the 1990 Census data for Oregon and other Northwest states. This has recently been expanded to cover all 50 states. Other demographic databases are Population Estimates for counties from 1990-1992 and USA Counties, a handy reference source with social, economic and governmental statistics spanning several years.

The economic databases include the 1992 Census of Agriculture, 1992 Economic Census, U.S. Imports and Exports, and the Regional Economic Information System (REIS) from the Bureau of Economic Analysis, which has income and employment data as well as short narrative summaries of regional economies. The recently added Consolidated Federal Funds Report details federal expenditures in all states, counties and municipalities.

Finally, the School District Data Book provides a wealth of demographic, financial and administrative data for all U.S. school districts. Users can find such things as the student-to-teacher ratio in their school districts and the level of funding by federal, state, and local governments.

Although these data bases are easy to use, providing access to them is labor intensive. All of the CDs contain files in dBase III format. By adapting DButil software developed at Lawrence Berkeley Laboratories, programmers on the project wrote software to read and format the data files. CGI scripts were written for the Web pages’ interactive forms. When a user queries the site, the programs extract the data directly from the CDs, which reside in drives attached to the Web server. One exception is the 1990 Census data. Because of the large number of CDs, these data files were extracted, subsets, and stored on the server’s hard drive. We have found that getting the data directly from the CD-ROM is almost as fast as accessing the dBase files directly from the hard disk.

Project staff members have taken care to make the Web site easy to use. Keyword searching is available within each database, allowing users to pinpoint statistics on specific topics. Besides the numerical data, the complete documentation from each CD was coded in HTML and is available on the site by clicking on the “info” buttons. Users can find out how the data was compiled, look up definitions of terms, and read about sources and authority of the data. There are also help buttons to provide context sensitive help in navigating and querying the data sets.

In a related project, the Government Information Sharing Project joined with the State Library and PORTALS to fund Jumplstart, a program that provides hardware, software, and training to small, rural school libraries and public libraries in Oregon that have limited access to the Internet. Twenty-four libraries were each given a 486 computer equipped with a printer and a 28,800 baud modem, and accounts were set up with local Internet service providers. Two representatives from each library came to Corvallis for a two-day training session on setting up their computers, navigating the World Wide Web using Netscape, and using the Internet for reference service. This program complemented the work completed on the Government Information Web site because it helped to ensure that libraries around the state would have access to the government information.

Although the initial goal of the project was to share OSU’s government information resources with other Oregon libraries, the Web has allowed us to reach people from far and near. Usage tracking software set up on the Web server even indicates usage by people around the world. Comments have been received from a variety of people, including a journalist in Ohio; government officials in Washington,
Change
(continued from page 14)

ing, how would they know where you stored the
important documents? If your colleague were hospi-
talized, would you know which file folders to turn to
for his or her part of tomorrow’s big meeting?
Just as each employee has his or her own filing sys-
tem at home and at work, each person who creates
a computerized file has an individualized system for
labeling and filing those on-line documents. In
libraries, we have come to rely upon such document
classification systems as the Dewey Decimal num-ering system and the Library of Congress call letter
system. But these systems clearly are not universal.
Some public libraries use LC and others use Dewey.
Many medical libraries use a modification of LC that
was devised by the National Library of Medicine.
Law libraries typically use yet another system. The
Universal Decimal Code was devised to be the
Esperanto of call numbering systems for libraries,
but it clearly is not as universally recognized as its
designers had hoped.

When we look at the array of files that are freely
available and accessible over the Internet, there is
nothing that could be comparable to a call numer-
ing system at the file-by-file level within any com-
puter. Most computers offer some sort of browsing
capabilities for the names of each file, but there is no
consistent system for browsing the contents of each
and every file that is accessible via the Internet.

But what about subject access to Internet files using
search engines such as Yahoo, which collates infor-
mation sources together from the World Wide Web?
I’m sorry to be the bearer of bad news, but despite
the decades of familiarity that librarians have with
systematic ways of describing the contents of books,
the lack of uniformity within our own libraries
(Sears, LCSH, MeSH, NAL, NLC and UKM) does not
establish a strong likelihood of classifying and categ-
orizing the contents of the World Wide Web.

Although many librarians are personally aware of
internationally established standards such as ISO’s
Z39.50 for interconnecting on-line library catalogs,
my research indicates that the actual number of
those standardized catalogs is very low. My experi-
ence of having searched over 400 separate libraries
using either NOTIS or INNOPAC brand software
indicates a lack of consistency even between cata-
logs running on the same brand of software. The
ability to customize features by turning on or off cer-
tain processing options has forced many library cat-
als to operate in a stripped down fashion when
connecting via a Z39.50 interface. Some library cata-
logs are designed to work with function keys, which
simply lose something in the translation when they
are connected to the Internet.

If we information specialists cannot get our own
computers synchronized with on-line catalogs in our
same cities, how can we expect individual’s at home
to overcome obstacles when they try to connect to
computers around the globe? The difficulties do not
stem from the computers themselves, but from the
lack of communication by the people who build
them, sell them, install them, and use them.

Alexander the Great built his library at a time when
communicating with people was much different than
today. But some things never change. The informa-
tion explosion that librarians are coping with pre-
dates both Marshall McLuhan’s writings, and the
Alexandria Library. Electronically, you can be linked
to people all around the globe in a flash. But it still
is a common language that separates England from
America. Just try using the Internet to find the offi-
cial colors of the Labor Party’s flag, and you will find
out what I mean.

Gary Klein has been using Internet resources at work
and at home since 1989. He has given conference
presentations at ALA, LOEX, NOTIS Users Group and
the Ohio chapter of ACRU, as well having journal
articles published on the lack of standards among
library OPAC systems and overcoming historical dif-
ficulties of problematic subject headings. Gary now
works for Willamette University’s Hatfield Library as
their Management & Business Economics Librarian.
You can reach Gary through the Web or by e-mail:
www.willamette.edu/~gklein

gklein@willamette.edu.

OSU Web Site
(continued from page 11)

D.C.; a city planner in Virginia; and students, faculty
and librarians from around the country.

Although it began as a demonstration project, with
an ending date of October 1996, the Government
Information Sharing Project will continue to maintain
the site. With new funding from PORTALS, OSU will
add several more databases and will experiment
with different types of CDs, including full-text. We
plan to work on improving the presentation and fea-
tures as well as adding to the content of the site,
developing a quality reference source on the Inter-
net for libraries to share.

Correction
In the last issue of OLA Quarterly, we
reported erroneously that Multnomah County
Central Library was originally funded by a
Carnegie grant. Although seven branches of
the Multnomah County Library were Carnegie
libraries, Central was not. The Library Associa-
tion of Portland purchased the land for the
Central Library, and Multnomah County levied
a tax for its construction. Thanks to June
Mikkelsen of Multnomah County Library for
pointing out this error.