

Collecting, cataloging and preserving digital publications

OCLC's Web Harvester helps the State Library to easily manage digital publications



Situation

Currently in state governments across the U.S., 80 percent or more of new government publications are created *only in digital formats* and published on the Web.

In Connecticut, the State Library is mandated by Section 11-9c of the Connecticut General Statutes to collect, preserve and make accessible all the official publications of Connecticut's state agencies. For over 150 years, the Connecticut State Library has collected, organized and preserved publications of state agencies. Now the State Library is mandated with having to preserve the digital publications of state agencies.

In order to fulfill this mandate, Stephen Slovasky, Unit Head of Bibliographic Services, Connecticut State Library, needed to maintain the documents in their original digital formats while easily integrating their acquisition and processing into existing library workflows.

"Anybody who knows anything about cataloging workflows knows that you want to keep things as simple as possible without oversimplification," Slovasky comments.

Solution

In order to meet all of the State Library's goals, Slovasky uses the OCLC Web Harvester™. The Web Harvester is integrated with both the OCLC Connexion® cataloging service and the CONTENTdm® Digital Collection Management Software. It enables the State Library staff to simultaneously catalog and capture state agencies' digital publications from the Web, then submit them to CONTENTdm. The integration with cataloging is vital to Slovasky's design.

"One of the reasons Web Harvester fits into our workflow so simply is that it is integrated with Connexion. And Connexion is the nexus of our cataloging workflow," explains Slovasky.

AT A GLANCE

The Connecticut State Library, Hartford, Connecticut

- Captured over 12,000 Web-published serial issues and monographs to date.
- <http://cslib.cdmhost.com/>

"One of the reasons Web Harvester fits into our workflow so simply is that it is integrated with Connexion. And Connexion is the nexus of our cataloging workflow."



In addition, the Web Harvester enabled the State Library to integrate its Web-based publications with subject access and record groups, as well as print and Web-based items.

Results

As a result of the State Library's established workflow, **the people of Connecticut and government staff are able to**

Continued

Collecting, cataloging and preserving digital publications

Continued



search for and access state agencies' digital publications via WorldCat®, the local catalog and the CONTENTdm end-user Web interface.

Aside from state agencies' Web-published documents, other digital documents include reports from the State Office of Legislative Research. Often these reports are issued several times a day when the Connecticut State Legislature is in session and several times a month when it's adjourned. These two- to four-page reports are HTML documents posted to a Web site. Using Web Harvester via the Connexion client, the State Library can catalog and then harvest documents directly from the Web site. Ultimately, the digital documents are sent to CONTENTdm and the OCLC® Digital Archive for access and long-term managed storage.

"It is our intent to keep these forever and to display them as originally published as far into the future as we need to," says Slovasky.

In addition to the ongoing work of harvesting Web-based publications, the State Library also has the option to use the Web Harvester to capture entire Web sites. For example, when a governor leaves office, State Library staff are able to capture the governor's entire Web site in order to ensure long-term management of the content.

To date, the Connecticut State Library has captured over 12,000 Web-published serial issues and monographs using OCLC's Web Harvester tools. Each catalog record includes a link to the digital item in CONTENTdm.

"To the end user, digital content management is secondary to having the item display as close as possible to how the original displayed," notes Slovasky, "and allowing whoever is displaying this document to decide how he or she is going to read it—print it, read it online or put it on their eReader."

For more information about OCLC's Digital Collection Services, visit www.oclc.org/services/collection/, call 1-800-848-5878, ext. 4301, or e-mail digitalcollections@oclc.org.