



# Library of Congress Digital Preservation Newsletter

## UNC-Chapel Hill Releases YouTube Archiving Tool

The University of North Carolina at Chapel Hill has developed a Web-based archiving tool as part of its NDIIPP [VidArch](#) project. The tool, ContextMiner, enables users to collect links to blogs and online videos along with extensive metadata.

A user begins by creating a scheduled, repeated collection activity called a “campaign.” ContextMiner has over 20 optional descriptive and administrative metadata fields for each campaign. The user then selects the content source from which ContextMiner will extract the data. Currently the options are YouTube and blogs but UNC plans to expand content options in the future.

ContextMiner displays the query results in tabled records. Each record contains hyperlinks: embedded links to YouTube videos, links to blog pages and links to related Web sites.

ContextMiner does not download and archive videos or blog pages; it only links to the Web source. The developers hope to eventually offer tools and policies for exporting and sharing videos, blog pages and metadata. More information is available at <http://www.contextminer.org/>.

## Digital Preservation Pioneer: Larry Carver



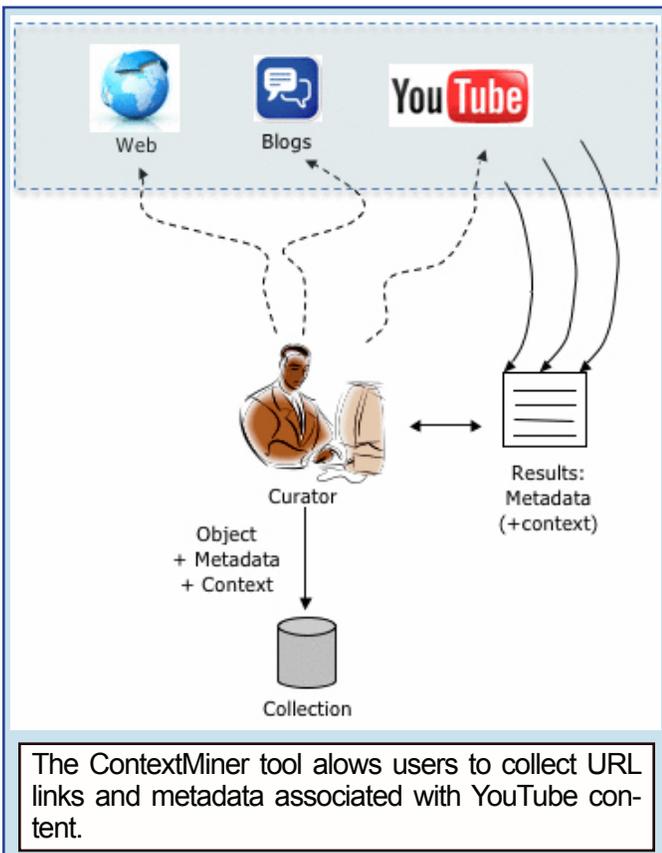
On November, 3, 2008, election night, almost every television news station and Web news site displayed an interactive map of the United States with real-time voting results.

The maps shared a common feature: geospatial technology analyzing a range of geographic-based coordinates and data. The technology has evolved over the last three decades from specialized academic obscurity to just another information-age appliance. Larry Carver is among the visionaries who enabled this transformation.

Carver began his career at the library at the University of California at Santa Barbara, where he helped build an impressive collection of maps, aerial photography and satellite imagery. His efforts resulted in the development of the Map and Imagery Laboratory in 1972.

As the MIL collections grew, Carver felt that geospatial data presented a unique challenge to the library. He believed that coordinate-based collections should be managed differently than book-based collections. “It became apparent that handling traditional geospatial content in a typical library context was just not satisfactory and another means to control that data was important.”

Carver and others spent years developing an exhaustive set of requirements for building a geospatial



Next the user creates keyword queries. For example, if the collection centered on Louis Armstrong the user might ask ContextMiner to query YouTube and blogs for the keywords “Louis Armstrong,” “New Orleans trumpet” and “Satchmo.” The user schedules ContextMiner to query sources daily, weekly or monthly. ContextMiner can also download contextual information such as tags, Web links and numbers of views.



information management system. The system had a number of innovative ideas. “We included traditional methods of handling metadata but also wanted to search by location on the Earth’s surface,” Carver said. “The idea was that if you point to a place on the Earth you could ask the question, ‘What information do you have about that space?’ as opposed to a traditional way of having to know ahead of time who wrote about it.”

An opportunity to develop that system arrived in 1994 when the National Science Foundation funded Carver and his team to build the Alexandria Digital Library. “Our concentration was to be able to develop a system that could search millions of records with latitude and longitude coordinates and present those results via the Internet.”

In the process of system development, Carver accumulated a great deal of national spatial content, much of which came from government agencies and private corporations. The geospatial collections are not limited to images. The ADL system enables searches for any data that includes a coordinate system, such as demographic data, census and socio-economic data, real estate and even biological systems. The possibilities are almost limitless; any information that can be related to a spot on the Earth can be searched.

Awareness of the value of geospatial data has been slowly spreading beyond the scientific community for the past decade. Businesses are especially interested in data that displays comparative changes over time, and Carver figured out a way to meet these needs through a “fee for service” arrangement. He said, “A lot of the materials that we collected were unique and they got the interest of environmental companies, lawyers and just about anybody who was trying to manage land resources. So we set up a model for charging for access to those collections.”

Long-term, reliable access to the geospatial data in the ADL system is essential to maintain services and the archives. In 2004, UCSB became an NDIIPP partner along with Stanford University in the National Geospatial Digital Archive Project and later, in a second phase, with the University of Tennessee at Knoxville and Vanderbilt University. This latter work related to development of the Federated Archive Cyberinfrastructure Testbed.

“FACIT creates a whole series of distributed nodes that know about each other,” said Carver. “And

as information is added in one node, any one archives can decide how many replicas it wants. That has two advantages. It provides an almost fool-proof backup system for your own archive and it also enables faster searching through distribution of data copies.”

The basic concepts behind ADL have been widely adopted by Google Earth, Wikipedia and others. And Carver could not be more delighted.

“I think it’s wonderful,” he said. “We weren’t trying to be the only game in town; we were just trying to raise consciousness way back in the early 1980s that this was a viable way of handling geospatial material. This approach lets people interact with data in a realistic way without having a great deal of knowledge about an individual object. It was a new way of dealing with massive amounts of information in an environment that made finding and accessing information much easier.”•

## Recent Meetings

- Partners in the Data Preservation Alliance plan for the future. [http://www.digitalpreservation.gov/news/2008/20081125news\\_article\\_datapass\\_meeting.html](http://www.digitalpreservation.gov/news/2008/20081125news_article_datapass_meeting.html)
- Over 50 delegates from countries throughout the European Union attended the third annual WePreserve digital preservation conference. [http://www.digitalpreservation.gov/news/2008/20081201news\\_article\\_wepreserve\\_meeting.html](http://www.digitalpreservation.gov/news/2008/20081201news_article_wepreserve_meeting.html)
- GeoMAPP partners gathered to identify strategies and solutions for the preservation of geospatial information at the state and local level. [http://www.digitalpreservation.gov/news/2008/20081118news\\_article\\_geomapp\\_meeting.html](http://www.digitalpreservation.gov/news/2008/20081118news_article_geomapp_meeting.html)
- Aarhus, Denmark became a hotbed of Web archiving during the European Conference on Research and Advanced Technology for Digital Libraries. [http://www.digitalpreservation.gov/news/2008/20081201news\\_article\\_archivinginaarhus.html](http://www.digitalpreservation.gov/news/2008/20081201news_article_archivinginaarhus.html)

## Announcement

The deadline for submitting presentation abstracts for Archiving 2009 to be held May 4-7, 2009 in Arlington, VA., is December 21, 2008. A PDF of the Call for Papers can be found at: <http://www.imaging.org/conferences/archiving2009>

To subscribe to this newsletter, go to [https://service.govdelivery.com/service/multi\\_subscribe.html?code=USLOC&origin=http://www.loc.gov](https://service.govdelivery.com/service/multi_subscribe.html?code=USLOC&origin=http://www.loc.gov) type in your e-mail address, scroll down and click on “Digital Preservation.” Past newsletters are archived in PDF/A at: <http://www.digitalpreservation.gov/news/archive.html>