National Digital Newspaper Program (NDNP)

The National Digital Newspaper Program (NDNP) [1] is a partnership between the National Endowment for the Humanities (NEH) and the Library of Congress (LC) to create and maintain an Internet-based, freely-accessible, searchable database of U.S. newspaper information and select digitized pages. This database is known as Chronicling America [2].

NEH provides 2-year grants for state institutions (awards) to digitize 100,000 pages of microfilm newsprint, published between 1836 and 1922. To date, 28 states have contributed to the program.

Acknowledgements

The National Digital Newspaper Program is funded by the National Endowment for the Humanities and supported by the Library of Congress. Chronicling America, including its API and the LC Newspaper Viewer software, has been developed by the Repository Development Center (RDC) at LC. Special acknowledgements include: Leslie Johnston and the staff of the RDC; Teri Sierra; Mark Sweeney; and Deborah Thomas.

Chronicling America

Chronicling America is the website providing access to the NDNP resources, which include:

- 140,114 newspaper records
- 385 title essays
- Almost 5 million digitized pages

Chronicling America is a partnership between the Library of Congress, the National Endowment for the Humanities and Library Services and Technology, the National Park Service, and the National Center for History and Culture.

API

Chronicling America provides an API to access public web protocols and data. The Application Programming Interface (API) for Chronicling America promotes a wide range of data use. Through the API, several views of the digitized content and metadata are publicly visible, with no restrictions to access. No API key is required [7].

OpenSearch

Searching across the Chronicling America Newspaper Directory of over 140,000 newspapers MARC title records is possible using the OpenSearch protocol and the OpenSearch Description document describing Chronicling America’s search engine.

CORS and JSONP Support

Chronicling America has been constructed to integrate with third-party JavaScript applications, with support for both Cross-Origin Resource Sharing (CORS) and JavaScript Object Notation with Padding (JSONP) responses.

Sitemap Protocol

Chronicling America follows the Sitemap protocol [9], providing an URLS file linked from the site’s robots.txt file.

Traffic

Opening Chronicling America’s data to internet search engines and commercial users greatly increased the project’s visibility and site traffic. In June 2008, the project reached a milestone of one million pages of content, and a site redesign added API access and crawler support. Site traffic doubled overall, and search requests multiplied fifty-fold. A May 2011 redesign added social media sharing tools, increasing the direct linking capabilities of the site. Chronicling America has also been harvested by commercial users and integrated into proprietary databases. One subscription genealogy site now refers about 5% of all total traffic. As detailed in Figures 6 and 7, use of the site continues to increase. From March 2011 to March 2012, total hits to the site increased over 100%. Genealogy sites and search engines that heavily use our API are also among the top referring domains.

References


Using NDNP Data

Chronicling America has welcomed academic research by providing API documentation on the site as well as joining the list of repositories for the NEH sponsored Digging into Data Challenge [12]. For a 2011 Challenge award, researchers from the Virginia Polytechnic Institute and State University of Toronto will mine the OCR text to examine public opinion and information dissemination during the 1918 Flu Pandemic. In a visualization project, Stanford researchers extracted records from the newspaper directory to map the growth of American newspapers [13] (Figure 7). Chronic History America also inspired a visualization project on the evolution of individual newspaper titles through mergers and acquisitions [14]. Link-upedia, a project developed by Ed Summers, shows how Chronicling America content has been reused or linked within Wikipedia [15]. Chronicling America’s new CORL text has been used in numerous other academic research endeavors.

References and Links


