FEBRUARY 4, 2015

MANAGE YOUR ACCOUNT   CONTAC

Crowdfunding Access to Archives | Backtalk

By Andrée J. Rathemacher on   February 2, 2015*(2015-02-02T11:55:22+00:00)*

Ryan Cordell, Northeastern University (NU), Boston, and his colleagues are studying how information went "viral" in 19th-century America, when newspapers and periodicals published short works of fiction, poetry, and other prose. Before modern copyright law, it was common for editors to reprint these texts, originally published elsewhere. The texts moved around the country through this network, resulting in a shared print culture. Cordell's research seeks to identify these shared texts, to examine which were reprinted and why, and to map how they traveled and changed as they passed from publication to publication.

To begin, Cordell and David Smith, an NU professor of computer and information science, downloaded the text of the entire corpus of pre-1860 newspapers available from the Library of Congress's Chronicling America*<http://chroniclingamerica.loc.gov/>* website. They performed a computational analysis using algorithms they developed to identify matching texts. Thus far, the project team has identified thousands of viral texts, including minor pieces by major authors that were more influential than previously recognized. They have also mashed up the data with other open data to reveal connections between viral texts and the expansion of railroads, the establishment of political boundaries, and local population characteristics.

BEHIND THE PAYWALL

Yet, according to Cordell, there are "glaring holes" in his research. His data includes no content from Massachusetts—and Boston was a major publishing center of the time. There is also very little available to him from New York or Philadelphia, also vital to the period. He lacks this content because it is locked up in commercial databases. Although his institution subscribes to a number of these databases, the ability to download the text for analysis is unavailable. Cordell has approached two of the publishers of these databases to negotiate such access, but the companies are understandably reluctant to release full-text content, as this would essentially mean relinquishing control over their intellectual property. In any event, it is likely that any permission to text-mine content would be granted on a case-by-case basis at an additional cost.

Before the digital era, libraries and cultural institutions entered into agreements with vendors to microfilm their collections and market the microfilm to other libraries, enabling broad access to the content that was not previously possible. As these vendors now build digital archival products, they have returned to the libraries to negotiate rights to resell the content in digital form.

Still, it is not only these legacy microfilm collections that are being marketed to libraries digitally. A perusal of the websites of Adam Matthew, Alexander Street, EBSCO, Gale, ProQuest, and Readex shows that they are working with libraries and other institutions to create newly digitized archival collections at a prolific rate.

For scholars, the problem with this model of digitization is that the materials are available only to subscribers, and text mining and other innovative reuses of the content are possible only by special arrangement under limited conditions.

That said, digitization is expensive, and some would argue that without these commercial partnerships mass digitization of unique archival materials would be impossible.

ANOTHER MODEL

There is an alternative for libraries that want to digitize archival collections and make them freely available to all. There have been a number of recent initiatives that use crowdfunding to cover the costs of making materials available via open access, for example, SCOAP3 for journals in high-energy physics, Knowledge Unlatched for scholarly ebooks in the humanities and social sciences, and Open Library of Humanities for articles and books in the humanities. For the digitization of archival materials, a similar initiative is Reveal Digital.

Founded in 2011, Reveal Digital relies on libraries to crowdfund specific digitization projects. Its initial effort, Independent Voices, will result in the digitization of over one million pages of alternative press periodicals in the collections of partner libraries. Digitized materials will initially be available only to libraries providing financial support for the project, but after an embargo period, the content will become fully open access and the full-text corpus will be available for data mining. Because of Reveal Digital's cost-recovery model, the contribution required by supporting libraries is only about 20 percent of the purchase price of similar collections offered through traditional publishing.

We should work to develop additional opportunities for libraries to fund open access digitization projects collectively. We need more experiments in

developing sustainable, alternative business models for making valuable historical primary sources, many of which are in the public domain, freely available to all. Librarians need to resist the enclosure of the cultural commons that is the inevitable outcome of the traditional publication model. This is especially important as new forms of scholarship emerge that require full access to digitized content.

Andrée J. Rathemacher is Head, Acquisitions, University of Rhode Island, Kingston. We welcome opinion pieces for BackTalk. Please send them to mschwartz@mediasourceinc.com

This article was published in Library Journal*'s February 1, 2015 issue. Subscribe today<http://lj.libraryjournal.com/subscribe/> and save up to 35% off the regular subscription rate.*

Filed Under: Backtalk*<http://lj.libraryjournal.com/category/opinion/backtalk/>*, Digital Content*<http://lj.libraryjournal.com/category/digital-content/>*, Digital Resources*<http://lj.libraryjournal.com/category/digital-resources/>*, LJ in Print*<http://lj.libraryjournal.com/category/lj-in-print/>*   Tagged With: archives*<http://lj.libraryjournal.com/tag/archives/>*, digitization*<http://lj.libraryjournal.com/tag/digitization/>*, LJ_2015_Feb_01*<http://lj.libraryjournal.com/tag/lj_2015_feb_01/>*   **Discussion: Leave a Comment***<http://lj.libraryjournal.com/2015/02/opinion/backtalk/crowdfunding-access-to-archives-backtalk/#respond>*

(PrintFriendly)   (Email)   (Twitter)

(Facebook)   (Google+)   (Tumblr)   (Reddit)

Share / Save

Comment Policy:
1. Be respectful, and do not attack the author or other commenters. Take on the idea, not the messenger.
2. Don't use obscene, profane, or vulgar language.
3. Stay on point. Comments that stray from the topic at hand may be deleted.

We are not able to monitor every comment that comes through (though some comments with links to multiple URLs are held for spam-check moderation by the system). If you see something objectionable, please let us know*<mailto:support@mediasourceinc.com>*. Once a comment has been flagged, a staff member will investigate.

We accept clean XHTML in comments, but don't overdo it and please limit the number of links submitted in your comment. For more info, see the full Terms of Use*<http://lj.libraryjournal.com/about/terms-of-use/>*.