

2015

Crowdfunding Access to Archives

Andrée J. Rathemacher
University of Rhode Island, andree@uri.edu

Follow this and additional works at: https://digitalcommons.uri.edu/lib_ts_pubs

 Part of the [Library and Information Science Commons](#)

Terms of Use

All rights reserved under copyright.

Citation/Publisher Attribution

Rathemacher, Andrée J. "Crowdfunding Access to Archives." *Library Journal* 140, no. 2 (1 February 2015): 43.

This Article is brought to you by the University of Rhode Island. It has been accepted for inclusion in Technical Services Department Faculty Publications by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons-group@uri.edu. For permission to reuse copyrighted content, contact the author directly.

Andrée J. Rathemacher

Crowd-funding Open Access to Digital Archives: In Support of New Models

Ryan Cordell and his colleagues at Northeastern University are studying how information went “viral” in nineteenth century America. In addition to news items, nineteenth century newspapers and periodicals published short works of fiction, poetry, and other assorted prose. Before modern copyright law, it was common for editors to reprint these texts, originally published elsewhere, for their own readers. In this way, the texts moved around the country through a network of newspapers and magazines, resulting in a shared print culture. Cordell’s research seeks to identify these shared texts, to examine which texts were reprinted and why, and to map how they traveled and changed as they were passed from publication to publication.

To conduct this research, Cordell and David Smith, a 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* website. Then they performed a computational analysis of the corpus 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.

Yet, according to Cordell, there are “glaring holes” in his research. His data includes no content at all from Massachusetts—and therefore Boston, a major publishing center of the time. There is also very little available to him from New York or Philadelphia, also important centers of nineteenth century publishing. 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 not available. Cordell has approached two of the publishers of these databases to negotiate such access, but the companies are understandably reluctant to release the full text content of their databases, 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 additional cost.

The reason the content Cordell needs is available only from commercial vendors dates to before the digital era. Libraries and cultural institutions entered into agreements with vendors to microfilm their collections and market the microfilm to libraries, enabling broad access to the content that was not previously possible. Now as these vendors seek to build digital archival products, they have returned to the libraries to negotiate rights to resell the content in digital form.

But 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 Press, 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 re-uses of the content are possible only by special arrangement under limited conditions. That being said, digitization is expensive, and some would argue that without these commercial partnerships to provide access to financial and technological resources, the mass-digitization of unique archival materials would not be possible.

But is there an alternative, given the limits of grant-funding and institutional support, 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 a library crowd-funding model to cover the costs of making materials available open access; examples include SCOAP3 for journals in the field of high-energy physics, Knowledge Unlatched for scholarly e-books 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 crowd-fund specific digitization projects. Their initial project, 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 strict cost-recovery model, the contribution required by supporting libraries is only about 20% the purchase price of similar collections offered under traditional publishing models.

As the crowd-funding model for all types of open-access content grows, we should work to develop additional opportunities for libraries to collectively fund open access digitization projects. 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 that turns our shared cultural heritage into privatized corporate assets. As we can see in Cordell's work, this is especially important as new forms of scholarship emerge that require full access to digitized content.

Sources:

Ryan Cordell (assistant professor, English Department, Northeastern University), in discussion with the author, January 2014.

Reveal Digital, "Independent Voices: An Open Access Collection of an Alternative Press," Reveal Digital, <http://www.revealdigital.com/voices> (accessed January 2, 2014).

Jeff Moyer (founder, Reveal Digital), in discussion with the author, January 2014.

Reveal Digital, Independent Voices project prospectus, <http://www.revealdigital.com/fileasset/Alternative-Press-Prospectus-121212.pdf> (accessed January 2, 2014).

Jeff Moyer (founder, Reveal Digital), in discussion with the author, January 2014.

Andrée J. Rathemacher is head, acquisitions at the University of Rhode Island, e-mail: andree@uri.edu.