BACKTALK

Crowdfunding Access to Archives
By André J. Rathemacher

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 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 who 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 pos-

We need more experiments in developing sustainable, alternative business models

sible. 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 crowd fund 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.

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

WWW.LIBRARYJOURNAL.COM REVIEWS, NEWS, AND MORE

FEBRUARY 1, 2015 | LIBRARY JOURNAL | 43
feedback

"If a student can get the first two years done at a community college and then go to a four-year institution to finish the bachelor's degree, that makes it cheaper and less debt for the student"

To cut crushing debt
I think one of the purposes of America's College Promise (Cathryn Kaufman, "America's College Promise," BackTalk, Lj 2/15/15, p. 40) is to try to cut down on some of that crushing student debt. If a student can get the first two years done at a community college and then go to a four-year institution to finish the bachelor's degree, that makes it cheaper and less debt for the student. Another purpose of the program is to try to address some of the shortcomings in K-12 [curricula]. Higher education is still voluntary. So, if a student had issues in high school, no matter what the issues were, community college can be a place to learn and address any deficits before they hit the job market.

—Katy Walker, Reference/Instruction Libr., Colorado Mountain Coll., Edwards

Lowered expectations?
I perceived President Obama's community college remark as a chance for those baby boomers who are losing their positions due to downsizing (Cathryn Kaufman, "America's College Promise," BackTalk, Lj 2/15/15, p. 40). Employers are no longer willing to pay for their retirement as well as their salary because they probably could pay someone cheaper who just graduated. A "free" education enables them to take classes for a more updated position, while not worrying about using money from their savings while they are laid off. Details have not been released about any stipulations if there are any. We should be focused more on those who are already involved with five to six figures (or more) of student debt onto which they may tack more debt when buying a house or a car or starting a family. I thought President Obama was going to take on this challenge, but maybe I was mistaken or he lowered his expectations to just the community college level.

—Jeanine Osterh, Info Specialist, Russell Branch, Chesapeake P.L., VA

Going nowhere
It all sounds very exciting (Cathryn Kaufman, "America's College Promise," BackTalk, Lj 2/15/15, p. 40), but this is going nowhere in the current Congress. The only thing addressing the issue during the State of the Union did is put the conversation on the table, where it could be for years, until we hit the next fiscal crisis.

—Name withheld upon request

A prepaid club
We don't use "free" in our library (Rebecca T. Miller, "Worth the Price: Reflecting on the Problem of 'Free,'" Editorial, Lj 2/1/15, p. 8). If I hear a staff member or member of the public call us "free," I politely correct them. We are not at point of service. We are a prepaid membership club. Our job is to provide as much value as possible for our members and potential members. This is paramount, especially because they so rarely feel as if they get a choice in paying dues.

—Spencer Smith, Dir., Little Elm P.L., TX

Amazed at libraries
I recently had lunch with an old friend, and at some point our conversation turned to public libraries (Cheryl LaGuardia, "Change for Researchers' Sake," Not Dead Yet, ow.ly/IUh4m). He had not set foot in one in over ten years and recently visited a Chicago Public Library branch. He was amazed at how "libraries have really kept up with the times" with the computer access, electronic resources and databases, and multiple forms of media available. While I was thrilled that he was so excited about the changes, internally I was screaming, "What did you expect?!" And, of course, making sure that it's not another ten years before he returns.

—Name withheld upon request

Do our jobs well
We librarians can go too far and push too hard in areas that make no sense to our customers (Steven Bell, "Trying Too Hard for Relevance," From the Bell Tower, ow.ly/IUhZ). Librarians need to get over patting themselves on the back at how great they are—how relevant they are—in providing "information" to our users. What "we" need to do is our jobs—and do them well—and the customers will give us value and priority that is most appropriate to them.

The library/librarian value proposition has changed: from keys to the kingdom to keys to all; from do it for you to self-service; from just in case to just in time; from containers and physical to information, not containers, not physical. The librarian challenge is to identify orthodoxies and propose change. We cannot be self-determined in our relevancy; that determination belongs to those who ask for our help.

—Name withheld upon request

Corrections
The "Editors' Spring Picks" feature (Lj 2/15/15, p. 30-35) contains an error: the hero's sister in Erika Swyler's The Book of Speculation (p. 33) is named Enola, not Flavia.

In "Crowdfunding Access to Archives" (BackTalk, Lj 2/1/15, p. 43), "the project team has identified thousands of viral texts, including minor pieces by major authors who were more influential than previously recognized" should have read "minor pieces by major authors that were more influential than previously recognized." Lj regrets the errors.