

## University of Rhode Island DigitalCommons@URI

---

Technical Services Faculty Presentations

Technical Services

---

2015

# Open Access to Archival Collections

Andrée Rathemacher

University of Rhode Island, [andree@uri.edu](mailto:andree@uri.edu)

Follow this and additional works at: [http://digitalcommons.uri.edu/lib\\_ts\\_presentations](http://digitalcommons.uri.edu/lib_ts_presentations)



Part of the [Archival Science Commons](#), and the [Scholarly Publishing Commons](#)

---

### Recommended Citation

Rathemacher, Andrée, "Open Access to Archival Collections" (2015). *Technical Services Faculty Presentations*. Paper 3.  
[http://digitalcommons.uri.edu/lib\\_ts\\_presentations/3](http://digitalcommons.uri.edu/lib_ts_presentations/3)[http://digitalcommons.uri.edu/lib\\_ts\\_presentations/3](http://digitalcommons.uri.edu/lib_ts_presentations/3)

This Article is brought to you for free and open access by the Technical Services at DigitalCommons@URI. It has been accepted for inclusion in Technical Services Faculty Presentations by an authorized administrator of DigitalCommons@URI. For more information, please contact [digitalcommons@etal.uri.edu](mailto:digitalcommons@etal.uri.edu).

**Boston-Area Open Access Advocacy Meetup**  
**Snell Library, Northeastern University, Boston, MA**

**July 14, 2015**

=====

**Brief Topic Talks**

=====

***Open Access to Archival Collections***

**Andrée Rathemacher, University of Rhode Island**

=====

**A Problem**

We have a problem.

There is a mad rush by library database vendors to digitize and bring to market new collections of archival materials.

Even a casual perusal of the websites of Adam Matthew, Alexander Street Press, EBSCO, Gale, ProQuest, and Readex shows that they are working with libraries and other cultural heritage institutions to create newly digitized archival collections at a prolific rate. A 2015 press release from ProQuest, for example, boasted that in 2014 the company digitized approximately 12 million pages of historical documents.

Why is this is problem?

It is a problem because our cultural heritage is being locked behind pay walls, available only to researchers affiliated with the libraries that can afford to purchase these collections. And, in most cases, the content is not readily available for innovative re-uses such as text mining.

Just as open access advocates argue against the norm of academics signing over copyright in their journal articles to publishers — so that the articles become the publishers' intellectual property, which they then monetize by restricting access — it is equally nonsensical that archival materials, much of which are in the public domain, should be similarly given away to corporate entities that wall them off to generate profit.

## **Research Needs Not Being Met**

Furthermore, even for libraries that can afford these collections, they aren't meeting the needs of our researchers.

As we know, OA advocates argue for open access to scholarly content not only because it is free to read, but also to reuse.

Some of you are likely familiar with the work of Ryan Cordell, a digital humanist here at Northeastern. He and his colleagues analyzed the full text of nineteenth century newspapers openly available through the Library of Congress's website *Chronicling America*. They created algorithms to track so-called "viral texts" that moved around the country as they were reprinted in multiple newspapers and periodicals. But, Cordell's research has "glaring holes" because his data include no content from Massachusetts and very little from New York or Philadelphia because this content is locked up in proprietary databases. Even though Northeastern subscribes to many of these databases, Cordell found that access to the full text for text mining was either not possible at all, or only through special arrangement, under limited conditions, for an added cost.

Similarly, Paul Fyfe at NC State who studies Victorian accidents and other topics from that era, had to press for a long time to get access to the full text in Gale's *19th Century British Newspapers* database. (In fact, it was his efforts in collaboration with Darby Orcutt of NCSU Libraries that was the impetus behind Gale's new license allowing for data mining and textual analysis... Though Fyfe found that this access involved multiple hard drives arriving in the mail with hundreds of folders of XML containing very bad quality OCR.)

## **One Solution**

I'd like to briefly present one solution to this problem.

Just as the Internet makes open access to published scholarship both possible and desirable, this is also true for archival collections.

And, as is the case even with open access systems of scholarly communication, there *are* costs involved.

Reveal Digital is a company that uses a library crowd-funding model to cover the costs of producing *open access* archival collections. In this sense, their projects are similar to

other OA initiatives like SCOAP3, Knowledge Unlatched, and Open Library of Humanities.

The company was founded by Jeff Moyer, a veteran of ProQuest, Gale, and National Archive Publishing Company. He learned from working with librarians on the EEBO Text Creation Project that libraries really care about making content openly accessible and will contribute financial resources to doing so. He also knew from experience that most archival collections achieve 90% of their sales within about 3 years after release, so other than the ability to value the digitized content as a corporate asset, it doesn't make much sense for these companies to claim ownership rights to it, since it ceases to generate revenue fairly early on.

Reveal Digital's first project is Independent Voices, which will result in the digitization of over one million pages of alternative press periodicals from in the collections of partner libraries. Digitized materials are initially available only to libraries providing financial support for the project, but after an embargo period, the entire collection will be openly accessible. Funding libraries also receive full support for mass text downloading. Reveal Digital makes no ownership claims to the digitized content and provides source libraries and rights holders with digital copies (images and metadata) of all the material they provide to a collection, plus the rights to make the material available online.

Other projects in the funding stage are:

- Student Non-Violent Coordinating Committee Digital Archive, 1960-1969
- Highlander Folk School Archives, 1932-1983
- Liberation News Service Archives, 1968-1981
- Sylvester Manor Archives, 1649-1996

Projects under consideration include:

- A jazz periodicals collection
- A student unrest collection that would focus on the Kent State shootings and other student protest movements at places like Michigan, Columbia, and Berkeley
- An anarchist collection
- A modern urban music scene periodicals collection

Reveal Digital operates on a strict cost-recovery, break-even pricing model. The costs for each project are shared transparently on the website, and include sourcing, scanning, and conversion; publishing and project management; editorial and rights clearance; royalties; systems and hosting; sales and marketing; and general administration.

Pricing for libraries is tiered by type of library and is based on an initial estimate of the number of libraries expected to support the collection. Libraries can expect to pay

about 20% of what they would pay for comparable collections from traditional publishers, as such projects set projected revenues at about five times costs.

Reveal Digital has established an Editorial Board consisting of academics and librarians who assist in the definition of collections and title selection. There is also a Steering Group that provides oversight on the cost-recovery model and advises on long-term archiving solutions and future directions. Libraries that support Reveal Digital projects have priority as future sources of material for new collections and input in defining future collection development.

According to Peggy Glahn, the company's Program Director, the ultimate vision is for Reveal Digital to be a facilitator and for libraries to be the real driving force behind the model. They are also exploring whether to pursue non-profit status to better be able to solicit support from granting agencies.

## **Conclusion**

I believe that we, as open access advocates, need to resist the enclosure of the cultural commons that is the inevitable outcome of the traditional business model that commercializes archival content. Instead we need to devote library resources to supporting experiments like this one that make archival content available to all and enable new forms of scholarship that require full access to digitized content.

## **Sources**

ProQuest digitization press release:

<http://www.proquest.com/about/news/2015/ProQuest-Moves-12-Million-Pages-of-History-From-Archive-Box-to-Search-Box-in-2014.html>

-----

Rathemacher, Andrée J. "Crowdfunding Access to Archives," *Library Journal* (February 2, 2015),

<http://lj.libraryjournal.com/2015/02/opinion/backtalk/crowdfunding-access-to-archives-backtalk/>

-----

Definition of OA: <http://legacy.earlham.edu/~peters/fos/overview.htm>

-----

Ryan Cordell:

Rathemacher, Andrée J. "Developing Issues in Licensing: Text Mining, MOOCs, and More." *Serials Review* 39, no. 3 (September 2013): 205-210.  
[http://works.bepress.com/andree\\_rathemacher/74/](http://works.bepress.com/andree_rathemacher/74/)

-----

Paul Fyfe on receiving and working with data from Gale:

Fyfe, Paul. "What is Fair Use of Seven Terabytes?" Presentation at ACRL/NEC Scholarly Communication Interest Group Spring Program, *Open Humanities and Digital Scholarship: Access, Innovation, and Support*, March 19, 2015, Northeastern University. <http://www.slideshare.net/paulcfyfe/what-is-fair-use-of-7tb>

-----

NC State University & Gale data mining agreement:

<http://www.infodocket.com/wp-content/uploads/2014/11/final-Gale-data-mining-press-release-1103142.pdf>

<http://www.prweb.com/releases/2014/11/prweb12331258.htm>

-----

Reveal Digital: <http://www.revealdigital.com/>

-----

Glahn, Peggy. "From Bombs to Bytes: The Evolution of Access Models for Cultural Resources." Presentation at ACRL/NEC Scholarly Communication Interest Group Spring Program, *Open Humanities and Digital Scholarship: Access, Innovation, and Support*, March 19, 2015, Northeastern University. <http://scig.acrlnec.org/media/32>

-----

Project Prospectus: Independent Voices: An Open Collection of an Alternative Press. December 12, 2012. Accessed July 10, 2015.

<http://voices.revealdigital.com/staticfiles/Alternative-Press-Prospectus-121212.pdf>

-----

Telephone conversation with Jeff Moyer, January 6, 2014.

-----

Herther, Nancy K. "ATG Original: Reveal Digital Unveils a New Business Model for 21st Century Archives." *Against the Grain*, November 14, 2014. <http://www.against-the-grain.com/2014/11/atg-original-reveal-digital-unveils-a-new-business-model-for-21st-century-archives/>

### **URI's support for Reveal Digital Projects (costs)**

- March 2014: Independent Voices, Cost: \$10,250.
- October 2014: SNCC Archive, Cost: \$4,000 (pledge).
- October 2014: Liberation News Service Archive, Cost: \$735 (pledge).

- October 2014: Highlander Folk School Archive, Cost: \$3,250 (pledge).