

The Other Side of the Computer: Spending a Summer with Digital Collections

DANIEL DAVIS

Utah State University

During the summer of 2010, I was granted a sabbatical from my day-to-day duties as the photograph curator in Special Collections and Archives. My main goal was to create a complete catalog with visual representations of all the photographs taken by Andrew Joseph Russell of the Union Pacific Railroad in 1868 and 1869. You've probably never heard of Andrew Joseph Russell, but I bet you've seen his classic images of the building of the first transcontinental railroad which are reproduced in textbooks, coffee-table books, and documentaries about the American West. Russell's photographs are scattered from coast to coast. Most of the original negatives are held at the Oakland Museum of California, with major collections at the Union Pacific Museum in Council Bluffs, Iowa, the New York Public Library, the Special Collections at the University of Iowa, the National Archives and Library of Congress, and the Beinecke Library at Yale. As well, private individuals own significant collections. My initial, and naïve, thought was to create a digital collection with links to digital representatives of all his images. In effect, I wanted to create a sort of uber-register spanning across multiple repositories with extended biographical and historical notes and comments on individual images. This concept could be an interesting paper in and of itself but suffice it to say that fiscal restraints and archival territorialism will not allow it to happen.

Introduction

Why do I feel a catalog of Russell's images is so important? Why not use the same five or six images that are trotted out over and over again? If I just needed the usual images like the usual researcher then I would have no complaints. Why this compulsion to see them all? Russell took nearly 1,000 images many of which (especially the stereo-views) have never been published. As we near 2019 and the 150th anniversary of the completion of the line, we should take a broader, more in-depth view of Russell and of the transcontinental railroad itself. The story of the great robber barons who squeezed huge profits from the line with their "creative" financing has been told. But the workers who lived and died to get it done remain mostly anonymous. As well, it is my contention that Russell's images were meant to be working documents for the Union Pacific. As such they can give us insight into how the line was built, and of the people who built it.

During my sabbatical, I spent my modest travel allowance quickly so I resolved to find what I could on the internet. Ironically, I went from a creator of digital exhibits to a user overnight. In fact, I had spent the previous 12 years creating digital exhibits. Now as someone on the other side of the computer, I was constantly surprised at how small details could create roadblocks for me. That said, I am old enough to be grateful for any Russell online images. His images are a hot commodity for collectors, and with the historical interest in building the first transcontinental

railroad, archivists have easily come to the conclusion that these types of images should be digitized.

So without further ado, here are my five big complaints. The first is that digital collections are deeply buried in a larger website. You try to find Russell's Civil War photographs at the Library of Congress, or his stereo-views in the online Central Pacific Photography Museum. I know you can search individual websites in Google's advanced features, but I think that less sophisticated users need good pathways and linkage to the actual material in the institutional holdings.

Secondly, different digital collections with images by Russell within the same institution have no cross-linkage. For example, at the New York Public Library there are no cross-links between a collection of stereo-views by Russell and the book "The Great West Illustrated," which features his large-format photos and his text. These cross-links should not be difficult to create, but they do require someone with historical knowledge and interest to insist on their placement.

Thirdly, often the verso side of an image is not available. Stereo-views often have additional information on the verso. I can tell for instance, if an image is by Russell even if that information is not included on the front by the design on the verso. Along with this complaint is the issue of low resolution (very low in some cases!) scans. I understand the desire of institutions to protect themselves against people who would "steal" such images and publish them without permission or attribution or without paying a fee, but in some cases the details of an image and the text are nearly unidentifiable.

Fourthly, often the name of the photographer is not included. For instance a landing page will mention Russell but then the individual metadata will not. Or sometimes the name of the photographer is listed as O.C. Smith when O.C. Smith was just a publisher who used Russell's images without attribution.

My final complaint is about the capricious and seemingly random nature of digitization. How is it that Frank Leslie's *Illustrated Newspaper*, probably the second most-read periodical of the 1860s and 1870s, is not digitized, but the *Corning Journal* from Upstate New York is? Copyright is not an issue, but perhaps this is a situation where the material is not unique to any one institution so that there is a reluctance to invest time and money on a generic product. As well, it is surprising that the two biggest collections of Russell images (both in museums) have only a handful of images online.

Is there a quick fix for these problems? Is this a matter of tweaking or should we look at it completely differently? The old approach to this problem would be to write a grant, convene a committee, stay in nice hotels in interesting cities, and adopt certain standards for the display of photographic images online. These standards might include a minimum of: front and verso displayed; a minimum 300 dpi resolution; title (if known); photographer (if known); links to images by the same photographer held by the same institution; and clear pathways from a home page to the individual digital object. While these are good suggestions, the adoption of such standards hinders and dissuades smaller institutions from posting images. Indeed our profession's fixation on standards has created a dichotomy of institutions with individuals brimming with confidence from their mastery of standards and acronyms. Individuals who have

achieved tenure or advanced in rank from this mastery, and who frequently make appearances in workshops and in articles. The other side, however, are individuals, usually from smaller institutions, who are racked with guilt and hesitation; afraid to do anything lest they be sneered at by the former group.

Almost by accident I stumbled upon another, perhaps more palatable, solution. As part of my “biographical note” on Russell I did research into his background. This research led me to one source which falls into another category completely. At first this website seems hopelessly flawed to the point that it simply isn’t useable. I’m speaking of fultonhistory.com which is a compilation of digital images and OCR’d newspapers from Upstate New York compiled by a very enthusiastic amateur historian who has a lot of time on his hands. As an aside, I should mention that Russell was from upstate New York and his local, gossipy hometown newspaper closely followed his western adventures. Where do you begin with this website? First off, the graphics are hideous. Secondly, the site says that 19,680,000 pages of newspapers can be searched, but the problem is that when you do a search it seems to return 19,680,000 hits. Thirdly, you can’t search by a single newspaper, you can’t search by date, you can’t browse page by page through a single edition, and in some cases you don’t even know what date the individual page is. Essentially each page was scanned and OCR’d as one PDF in a series based on the newspaper title and a range of years. When you bring up the PDF, you don’t know the date unless it’s printed on the newspaper but rather you know it’s from a two or three-year span (which is not nearly exact enough for picky researchers). When you type in “A.J. Russell” as an exact phrase you get 5,000 hits (the maximum it will bring up) and when you put in “Andrew Joseph Russell” as an exact phrase you get one hit.

After quitting the site in disgust I said I would give it one more chance, and I’m pleased I did because the second time around I found a wealth of previously unpublished material on Russell. I described the method I finally used to successfully navigate the website to a colleague thusly, “I hopped on one leg going in counterclockwise circles while patting my head.” One problem was that references to Russell might be under a phrase like, “the photographer Russell,” or “Captain Russell,” or “Capt. Russell” or “The Union Pacific photographer,” or “Russell’s images.” So I started to string together words like Nunda, and Russell, and Railroad, combined with words like photograph, stereograph, images, etc. and then added years. Of course, the OCR was inaccurate. I calculated that it only got about 1 in 3 words correct, but by putting a string of words together I was more likely to get a hit. Then I would create a big list and scan through individual listings and select those that I thought likely had something, doing a “Find in page” search on the actual PDF. Also the search display showed the text where Russell appears in the PDF, and I could tell at a glance if it was my Russell. Once I figured out that the file names, which it seemed at first were random numbers, had been sequentially scanned by date (rather than putting the actual dates in the file names), I was able to use a file name in the search box and figure out a date for those pages that didn’t include a date. Also, where I knew there was something about Russell on a certain date I could “guesstimate” about where it would be and type in different numbers until I found the right date. So if I eventually hit “pay-dirt” with this website, imagine how a better-designed one could help my research!

Ricky Erway wrote that, “Mostly we have been creating portals that lead to other portals that eventually lead to deep collections. Each collection has to be discovered and searched

individually. How many of them do how many users ever find?”¹ So even a site so highly flawed as fultonhistory.com can reveal hidden treasures because of its 19,000,000 PDFs. In fact I’ve had hits on this site while doing other Google searches related to my research (as a third aside, I should mention that I have never actually “googled” Russell’s name). Perhaps a better approach is to digitize “for the masses” and let the user sort it out for him or herself. The sites we’ve been creating are not really designed to be “googled,” but the research shows that the vast majority of patrons (some 80-90%) are finding our material not by navigating deeply around in our websites, but through search engines.

Now I am wondering whether the time-intensive, selective digital collections I used and created have inherent flaws that can readily be improved upon, or whether mass-digitization projects that stress quantity over quality provide a ready answer for both advanced users like myself as well as first-time and casual users. Maybe we can bring both types of digital material closer together? If both kinds of sites utilized historical expertise, bringing the content people together with the tool people, that would be a great start. It has been my experience that those who create digital collections frequently don’t use the collections or they don’t do broad-based research across a variety of institutions (or even sometimes don’t double-check the exhibits they create after uploading the content!). There seems to be a profound disconnect between “content” people and the “tool” people. Perhaps both mass-digital projects and more in-depth sites could be tweaked to provide services that cater to both first-time users and more advanced users such as myself.

I’m coming to a conclusion which has surprised even me. That with digital projects, quantity matters more than quality as long as you have broad search ability. Although I’ve preached digitizing only by specific selection, now I’m rethinking this approach. Let’s save the expense of writing a grant and convening a committee. For me there are three key components in a good digital project. First and foremost, get the material out there. Second, create a bunch of keywords. Third, consult with someone who has historical expertise in the area of the digital material and ask questions. Digital collections should be dynamic and ever-changing. If there’s something that is easy to change then do it, if it’s not so easy to change then, oh well, it will still be “out there.” For me it was better to have something up with a lot of keywords that is imperfect than a meticulous and impressive-looking digital exhibit with a lot of keywords. Quantity trumps quality. I’ll spend the time looking through the quantity to find the quality. Is not that sort of the idea with archives in the first place?

References

Erway, Ricky. “Supply and Demand: Special Collections and Digitisation,” *Liber Quarterly*, 18 (3/4), December 2008.

ARTstor, ARTstor Slide Gallery, Russell, Andrew J.
(<http://www.artstor.org.dist.lib.usu.edu/index.shtml>)

¹ Erway, Ricky. “Supply and Demand: Special Collections and Digitisation,” *Liber Quarterly*, 18 (3/4), December 2008.

Boston Public Library, Print Department, Western Photographs
(<http://www.bpl.org/research/print/western.htm>)

Central Pacific Railroad Photographic History Museum Archive, Andrew J. Russell Stereograph Catalog (http://cpr.org/Museum/Russell_Catalog.html)

Denver Public Library, Western History and Genealogy, Digital Collections
(<http://digital.denverlibrary.org>)

George Eastman House, International Museum of Photography and Film, Still Photograph Archive, Clarence King Survey of the 40th Parallel (<http://www.eastmanhouse.org>)

Library of Congress, Prints and Photographs Division, Civil War Glass Negatives and Related Prints (<http://www.loc.gov/pictures/collection/cwp>)

New York Public Library, Digital Gallery, Robert N. Dennis Collection of Stereoscopic Views.
(http://digitalgallery.nypl.org/nypldigital/explore/dgexplore.cfm?col_id=361)

Oakland Museum of California, The Andrew J. Russell Collection.
(<http://museumca.org/collection/the-andrew-j-russell-collection>)

The Union Pacific Railroad Museum, Council Bluffs, Iowa, The Union Pacific Collection, Photography Archives, A.J. Russell Views.
(http://www.uprr.com/aboutup/photos/1860s_1870s_ajr.shtml)

University of Iowa Libraries, Special Collections & University Archives, MsC 159, Levi O. Leonard Railroadiana.
(<http://digital.lib.uiowa.edu/leonard/>)

U.S. National Archives and Records Administration, College Park, MD, Still Picture Records Section, Special Media Archives Services Division, Record Group 77, Records of the Office of the Chief of Engineers, Timothy O'Sullivan images
(<http://research.archives.gov/description/519442>)

U.S. Geological Survey, Photographic Library, A.J. Russell Collection
(http://libraryphoto.cr.usgs.gov/cgi-bin/search.cgi?search_mode=noPunct&selection=Russell,%20A.J.|Russell|A.J.@)

Utah State History, Research Center of Utah State Archives and Utah State History, Classified Photograph Collection
(http://history.utah.gov/research_and_collections/photos/index.html)

Utah State University, Merrill-Cazier Library, Special Collections and Archives, P0349 19th Century Western Stereo-views Collection.
(<http://uda-db.orbiscascade.org/findaid/ark:/80444/xv28820>)

_____. Merrill-Cazier Library, Special Collections and Archives, P0019 *Sun Pictures of Rocky Mountain Scenery* Collection.
(<http://uda-db.orbiscascade.org/findaid/ark:/80444/xv36932>)

Yale University, Beinecke Rare Book & Manuscript Library, WA Photos 399
(<http://beinecke.library.yale.edu/digitallibrary>)

_____. Beinecke Rare Book & Manuscript Library, ZZc10 864un, A.J. Russell Photographs of the Construction of the Union Pacific Railroad
(<http://beinecke.library.yale.edu/digitallibrary/russell.html>)