“Looking for a radically open digital landscape”
presented on June 23, 2016 at the RBMS conference
Sarah Werner (sarah.werner@gmail.com)

If rare books and manuscript libraries used to be places where selected people were invited in to witness the display of special items, and those with sufficient expertise were allowed to use items under careful supervision, they now can become radically more open. More institutions are letting beginning scholars—even undergraduates or maybe younger students—have access to their physical collections, more libraries are placing digital facsimiles online for viewing to anyone who comes across them, and more are reaching out to audiences instead of hoping that they will stumble across them. It’s a shift from modes of limited access, expert authority, and control to ones of openness and sharing. It is a potentially radical way of changing how we think about what we do as special collections librarians.

Why do I emphasize potentially? In this digital landscape, we can open our collections with nearly unlimited numbers of people, who can then share their knowledge with untold others. This sounds like a recipe for openness, doesn’t it? For a radical reimagining of what special collections are and who our audiences are?

Before looking at how we experience and create this seemingly radical digital landscape, let’s review the analog one from the perspectives of our users.

For the uninitiated—the new scholars, the young student, the member of the general public—rare book and manuscript libraries are mysterious places behind locked doors. They are often hidden up on the top floors of libraries or behind closed doors that you don’t normally venture past. They have security guards or other staff sitting at a desk at the front looking at you before you can go into the reading room. You need to apply for permission to look at their collections, sometimes just showing your ID, but often providing letters of reference and a statement of your plans for research.

Once you’ve made it inside there are myriad procedures and more locked doors. You can’t bring in bags, or pens, or drinks, or cough drops. You might have to sign in and sign out. You can’t go into the stacks, browsing for books, but you have to use a catalog or a finding aid. Maybe you’ve never used a card catalog before, so you need to get used to how that works. There could be an online catalog, which might be integrated into the main library catalog’s interface, so it could at least be familiar, although then you have to figure out how to filter out all the non-rare books and all the electronic copies. (Don’t even get me started on how finding aids are completely not intuitive to first-time users.) Finally you find the manuscript you want to look at, but then you have to figure out how to ask for it. Is there a call slip? Have you ever used a call slip before?

Then you wait. Someone brings your manuscript to you, or maybe you sit patiently and return to the desk to pick it up when it’s ready. There are special supports for the book so that it doesn’t break

Werner, RBMS16, p1
and special weights to keep it open. Maybe someone shows you how to use these, maybe you surreptitiously look around and copy what others are doing and hope no one notices you’re clueless. You love this manuscript! It’s the most interesting thing you’ve ever seen! You want to take a picture of it to share with your friends and your Aunt Matilda, who loves old bibles, but is that allowed? Are there signs telling you to do it or not to do it? Is anyone else doing it?

I could go on, but you see my point: special collections libraries are not like libraries that most people use, for reasons that are very good but also have the effect of sending the message that THEY ARE NOT FOR YOU. I should add, of course, that for those of us who make it through those hurdles and get to love special collections libraries, they are the best places on earth, populated by staff and researchers who are (usually) eager to help and to share in your enthusiasm.

So, what about the digital landscape of special collections? (A quick aside: what makes up the digital landscape is huge, and includes digital collections, web archiving, social media, born-digital materials. For time’s sake, I’ve limited my exploration here to digital image collections, but I’m happy to discuss those issues in the Q&A or after.) Is our digital landscape more welcoming, navigable, open than the analog one? In theory, of course! You don’t need to be an expert, you don’t need anyone’s permission to enter, no one needs to know what you’re looking at, all that information is there for your use. In practice? Well….

For one thing, it’s fragmented—it’s not that simple to enter. If you know where a book is held and you want to see if there’s an image of it, that shouldn’t be tough: you go to that library’s website, navigate to their digital collection, and search! But that’s harder than it sounds: it’s not always easy to find where a library’s digital collection is, or how to search within it, let alone to work out whether or not the images are licensed for what you want to do with it. Often there isn’t a digital image collection, but a series of digital exhibits, each of which has to be browsed separately. All that trickiness of searching and asking permission—those are often replicated in our digital spaces.

Looking for images of a known item in a known repository is the easiest scenario for a researcher. What happens if you’re looking for images of a specific work but you don’t know where you might find them? This is more of a nightmare. If you’re lucky, you’ll find what you need in one of the few aggregating sites like DPLA, Europeana, or Umbra. If you work in the early modern period, you could go to Early English Books Online—if your institution subscribes to it—or to the Universal Short Title Catalogue, which includes links to many open-access images. (The new English Short Title Catalogue, when it arrives, should be able to include more links to digitizations beyond EEBO and ECCO.) Odds are, if you’re looking for openly accessible images, you’ll probably end up googling what you want and keeping your fingers crossed it will show up. (If it’s a major author’s work, like, say, Shakespeare’s First Folio, you face better odds than if it’s not, so praise be to folks like Desiree Henderson, who maintains an index of digitized diaries that the rest of us can benefit from.)
I was curious about what it was like looking for images along these lines, so I thought I’d try it. I started with Margaret Cavendish’s *Blazing World*, a work first printed in 1666 with a second edition following closely upon its heels in 1668. Cavendish is an amazing woman and an important early modern figure; *Blazing World* is often described as the first work of science fiction, and it can be read as responding to and participating in the scientific conversations of its day. It seemed like a good candidate for digitization: it’s significant in its field, of interest to scholars and potentially to wider audiences. I started by looking up *Blazing World* on the ESTC, which actually took me a bit of time, because the 1666 text is published with her *Observations upon experimental philosophy* and so when I was browsing the short title catalog, where it appears under that title, I didn’t at first realize what it was. Eventually, though, I had the dates of the first editions, and I cheated by looking at the images in EEBO, so that I was thoroughly armed to recognize the books when I came across them. Alas, I did not come across them. I tried DPLA, Europeana, Google. I didn’t quite believe I couldn’t find it, actually. I knew there was a digital project to create an edition of *Blazing World* and so I looked at that, hoping it would provide a link to images. After half an hour, I decided it didn’t exist, and a couple of Cavendish scholars confirmed its absence.

Disheartened, I tried a few more works, trying to go for texts that were better known. *The Narrative of the Life of Frederick Douglass* I quickly found on Internet Archive, although I landed first on the 1852 edition and had to do a bit more research to realize there was an earlier 1848 one. I googled Christina Rossetti’s *Goblin Market* and found the British Library’s online essay about it, which led me to their digitization of the second edition.

Then, feeling a bit pleased with my success, I thought of John Bunyan’s *Pilgrim’s Progress*, first published in 1678 and republished in countless editions in more than 200 languages, and still in print today. Again, I wanted to ignore EEBO and find an open-access version. So I started with the aggregators: DPLA and Europeana. In both places I found many editions. Europeana was chock full of 19th-century versions, but I was defeated by an inability to figure out how to sort the results by publication date, so while I’m fairly sure there are no 17th-century editions there, I’m not positive. (Remember my caveats about the difficulty of interfaces in the analog landscape?) DPLA was easier for me to manipulate, and I found the wonderful plates illustrating the 1682 5th edition, digitized from New York Public Library’s collections, but the earliest complete edition I found was the 25th one from 1738 in a poorly done black-and-white digitization by Google. And remember—I’m doing these searches with some amount of expertise. What challenges would a general user face?

I know, these examples are heavy on literary works and early modern ones at that. It’s hardly representative of our full digital landscape. In my defense, it’s what I was trained in and I was trying to avoid the complications of looking for texts printed after 1923. (Copyright is the bane of a digitizer’s existence.) I encourage you to play this game with your own areas of interest, and we’ll see that the more diverse the pool of materials we’re looking for, the less likely we are to find them online.
Tim Hitchcock, an historian of 18th-century London, made this point recently when looking at the biases behind digitization selection of historical materials. Thanks to the canonical preferences of Google’s digitization program and those of the mid-20th-century microfilming projects, he writes, the selection of whose pasts to digitise has been profoundly conservative – giving new breath to old ideas and old texts. It is easier to find details of the Gordon Riots of 1780 online, than it is to locate the archives of the Mau Mau uprising and its suppression; easier to look up the financial records of 19th c. slave holders; than the records of Apartheid. Easier to trace a 19th-century history of glorified empire, than the messy details of decolonisation. (“Privatising the Digital Past”)

Digitization that we drive ourselves is not necessarily better—a 2014 study found that only 1/3 of cultural heritage institutions had a digitization strategy. (Enumerate, “Survey Report on Digitisation in European Cultural Heritage Institutions 2014”) They were imaging, alright, just haphazardly. Is that going to do a better job than the biases of past efforts? And if there are strategies, are they going to focus on an institution’s collections or on a field’s needs? Do we work in silos or do we work together?

I did one last search for something that I knew existed as a digital facsimile, something that fits right into the biases of imaging efforts: Shakespeare’s First Folio. Pretending I was a vaguely curious general reader, and not the obsessive woman who maintains a descriptive catalog of what is currently 16 digital copies of the book, I searched “Shakespeare First Folio,” went to Wikipedia, and then had the option of 5 different digital copies to choose from. Success!

But let’s pause for a moment to note that contrast: 16 digital facsimiles of the First Folio and not a single one of Blazing World.

Think of the resources that go into imagining 16 copies of the First Folio. 16 copies of approximately 900 pages each. 16 copies that do not differ from each other in particularly meaningful ways. I am well aware of the value of copy-specific traits and of the important work Charlton Hinman did in collating over 50 copies to learn how the First Folio was printed. But there is only one significant textual variant among First Folios and that’s how the presence of Troilus and Cressida is or is not signaled. And all 16 copies that are imaged show the third state of this variant, the one where the play is smoothly incorporated at the start of the tragedies. These 16 copies are, for the most part, very bland copies of the First Folio, with hardly anything distinguishing them from each other. Two of them—the Meisei and the recently found St Omer copies—have interesting marginalia, so there’s some value to those specific copies being imaged. One of them—Penn’s copy—is the granddaddy of them all, digitized in 1998, so that certainly made sense then to be done; Brandeis’s was imaged around the same time. Two of the Folger’s three publicly accessible copies were imaged when they were loaned for exhibitions, which is good for security, and once you’ve imaged it, you might as well share it. But the rest of the 16? 8 of them have been imaged in the last

Werner, RBMS16, p4
three years. The only exception to my frustration that I’ll grant is the Bodleian’s digitization, because they built such a wonderful interface for it, something leaps and bounds beyond what else is out there, something that is actually usable for researchers and general interest.

16 First Folios, and not a single *Blazing World*.

Now it’s possible someone could argue that Shakespeare is 16 times as important as *Blazing World* and therefore it makes sense that 13 different institutions would want to digitize copies of Shakespeare’s book, and not Cavendish’s, even though at least two of those libraries own Cavendish’s as well. If that’s the case, we would certainly expect that these digital copies would help us understand why they’re so important and why it was so necessary for these libraries to image them.

Well.

Many of the copies aren’t accompanied by even a scant few lines about the book, and the only library that provides more than a paragraph of context alongside the First Folio is Leeds. If we look further afield for context not directly attached to the digital facsimile but in approximately the same vicinity, we find (if we look hard enough) a link to a pdf of Peter Blayney’s unrivalled introduction to the book at Penn’s site. The Folger has, hidden on its website, some very brief and very basic information about what the book is, if not much of an argument about why it matters.

What sort of outreach is this? It’s the kind that assumes that users already know what they’re looking at or that they’re not interested in or capable of understanding what is in front of them. “Here, look, we own a First Folio! See? Shakespeare!”

Now, looking at pretty pictures is one of the types of outreach that libraries can do and it’s a perfectly valid thing to do. I like to browse for pretty pictures, and so do lots of librarians and researchers I know. And some of you will want to push back right now and say, Hey, that’s what digital collections platforms do—they provide for browsing and searching images, not reading essays. And that’s true. That is what the bulk of the platforms we use do. But if part of the mission of libraries is not just to collect and to preserve but to educate, how is what we are doing serving that mission in its entirety?

I know, digitization is already resource intensive. There’s the selection (assuming that you’re selecting things deliberately, rather than haphazardly), there’s conservation, imaging, processing, managing the platform. Now I want you to start writing essays to go with each of your items? Who has the time??

Let’s imagine that one way to fulfill your educational mission also helps you with your outreach: perhaps researchers and users could use your materials to create educational tools. Maybe they can

Werner, RBMS16, p5
curate exhibits, build online courses, create teaching modules that would help others learn from what you’ve shared. But let’s look, again, at whether our digital collection landscape allows for that creation.

At a minimum, as a user, what you need is a set of digital images that are downloadable and licensed for educational use. That would let you use them on a website with your students. But let’s say you want to write about that material on your blog. Is that for educational purposes, for private research? Maybe you need a non-commercial license. But let’s say that your personal website is connected to a book that you’ve published. Is that use for non-commercial publication?

Let’s imagine that what you want to do is remix these images. Perhaps you’re interested in miscellanies and you want to create a digital miscellany mixed in with your own materials. Are the images licensed to allow derivates? Maybe you’re an artist and you want to remix these works to make your own new work. Perhaps you then want to sell your work because you need to eat and to live and to fund new projects. Are they licensed for that? And with all of these projects, do you have access to the information and formats that you need to do your work? Are images only available as pdfs, one at a time, in low resolution? The technology barriers can range from inconvenient to insurmountable. (Most of those First Folios allow users to download jpegs, although only 4 have images greater than 2000 pixels, and none make it easy to batch download. Three-quarters of the 16 place some sort of restrictions on the images’ uses, ranging from not allowing any reuse at all to educational to share-alike; 3 are in the public domain and let users do anything they wish with them.)

I’m not going to dwell on the wrongness of trying to claim copyright over reproductions of public domain works. It’s very, very wrong. You’re on shaky legal ground and even shakier moral ground. If you need a refresher on why libraries should not be licensing public domain items, you should read Michelle Light’s 2014 RBMS talk, published in the March 2015 issue of RBM, and then you should share it with your colleagues.

I will add one thing about licensing: if you’re bringing in commercial companies to do your digitization for you, you are not adding to an open digital landscape. You are allowing access to a set of users who are fortunate enough to belong to institutions rich enough to pay for those subscriptions. Is that divide really furthering our mission?

Libraries are built by and for communities. Yesterday’s plenary speakers spoke powerfully about the importance of a full diversity of communities seeing themselves reflected in collections. That is at least as true for the digital landscape of special collections as it is for the analog. We need strategies to digitize our collections so that we don’t image the same work over and over again but so we can begin to scratch the surface of our rich heritages in books and manuscripts. We need to create contexts for those works so that images of all of our pasts continue to have meaning into our future. And we need to let people use those images as they see fit. We need to provide our expertise but
then we need to let users be experts, too. I cannot tell you what Shakespeare means to you. I cannot
tell you what Margaret Cavendish or Frederick Douglass or *Pilgrim’s Progress* mean to you.

The radical potential of digital tools for special collections is they let everyone use rare books and
manuscripts. They let everyone read them and destroy them and remake them and carry them into
the future. We haven’t reached that radical openness yet. But we can come closer to that possibility
by holding on to our roles as custodians and educators and by letting go of the need to own the
access to and uses of our images.

*[postscript: After I gave this talk, I learned that there was a copy of the 1668 *Blazing World* on Google
Books. My response to that is in a short blog post, “searching for a Blazing World.”]*