Plenary Sessions
Access to Freely Available Journal Articles: Gold, Green, and Rogue Open Access Across the Disciplines

Michael Levine-Clark, Dean of Libraries, University of Denver
John McDonald, Associate Dean for Collections, University of Southern California
Jason Price, Director of Licensing Operations, SCELC Library Consortium

The following is a transcript of a live presentation at the 2016 Charleston Conference.

Michael Levine-Clark: We’re exploring the access and discovery to freely available articles, and we’re deliberately looking at not just open access content but anything that is freely available to a user on the web. From a user perspective they might care philosophically whether it is open access versus something that they are getting pirated access to, but the reality is that they may often not even know which type of access it is. So, we’re looking at gold open access, green open access, and rogue and pirate open access, stuff that maybe you shouldn’t quite have access to.

The library, we know, for many users is not the starting point. A recent ITHAKA report, as well as the New Media Consortium Horizon Report, has talked about this issue that users start very often from Google, from Google Scholar. They don’t start from library sources. The ITHAKA report talks about the fact that while discovery services for students are often important, much more often they are starting their searches from other places from the open web. And we’ve got data that backs that up. This is referral data to a particular publisher (see Figure 1).

The pie chart is the University of Denver, my institution, and this is almost a year’s worth of data for a particular publisher, and this is to the licensed content that we have at the University of Denver. Thirty-nine percent of the referrals to our context, to this publisher’s content, came from our library discovery services. So, from the discovery service, from the resolver, from the catalog, from databases; so library tools broadly speaking. Sixty-one percent came from other places, right? So, 32% came from Google and Google Scholar together. Twenty-seven percent were not sure where it came from; there is no clear originating source. But the key there is that for users very often they’re getting to our content from sources that are not the library or not library-specific sources. The pie chart is equivalent to this particular bar chart on the graph, so these six bars are six different institutions, University of Denver is one of them. And the bold content at the top, or the bold sections at the top, are the library-originated referrals and you can see in the green, the blue, and the red at the bottom, the stuff that’s coming elsewhere. Most of these referrals at these six institutions are coming again from outside the library. They are not coming from library discovery services or the library catalog.

![Figure 1. Single publisher referring Site URL data.](image-url)
And people are getting to content in a lot of different ways. One of these ways is ResearchGate. ResearchGate, as most of us know, is a sort of a social research tool where people can post content, people can share content, and people can ask for contact. There is metadata about this particular article in ResearchGate, but there is also an icon where a user can request that full text. I’m a member of ResearchGate. Many of us are, and one of the sort of annoying features of ResearchGate is that you get a lot of e-mail from them asking you to post stuff, right? So I’ve got a bunch of notifications here from people who want me to post something. ResearchGate doesn’t actually tell, it doesn’t help you determine whether you have the rights, as an author, to post a particular article, and very often the things that get posted on ResearchGate are not versions of the article that should be made freely available. They are rogue open access.

And then there is Sci-Hub. Sci-Hub is a tool that is out there with articles that are pirated from all sorts of different sources. Sci-Hub has been quite in the news, including this really detailed study of usage and the history of it that was in Science magazine last year. One of the things that was really interesting about this study is the number of people coming from places where they’d have legitimate access. So, from institutional sites, they’re going to Sci-Hub even though they are at universities that have access to a lot of this content. And one of the things that is interesting is that they tell us they’re going there for convenience. So, the orange bar on the slide (see Figure 2), the 23% and the 17% there, the convenience factor, so a combined 40% of the users there say they come to Sci-Hub even though they may have access, right? So 51% say they come because they don’t have access. Seventeen percent say that they use Sci-Hub because it is more convenient than the library or other sources that they have access to. Twenty-three percent say they object to the profits of publishers. That 40% probably has access, but they are choosing to use Sci-Hub anyway, and this is of 11,000 researchers, this survey. Eighty-eight percent of those surveyed said that they don’t actually believe that it is wrong to download pirated papers, so that is an issue that we should all be concerned about, right? That they are using Sci-Hub, and they don’t care that it’s pirated. They’re using Sci-Hub even though they probably have access in other ways.

A recent study shows that in 2013 we actually passed the 50% point for open access content on the web. In April of 2013, 50% of the peer-reviewed articles that had been published in 2011 were available in some form of open access, green or gold, on the web. So, we decided to investigate sort of the broad availability: green, gold, rogue, and pirate, pirated meaning on Sci-Hub, of freely available article content. We randomly selected 300 articles that were indexed in Scopus and published in 2015. A hundred of them are from the arts and humanities, and a hundred of them are from the social sciences, and a hundred are from the life sciences, and all of them, again, randomly selected.

We’ll be talking about a few definitions sort of as we go through. I want to just be clear what we mean by these things, by these terms. Availability means the presence of full text in a free version. Right? That we found some full text freely available on the web. We didn’t have to login in any way. We searched in four different locations. Our search locations were Google Scholar, Google, ResearchGate, and Sci-Hub. Again, two open sources: ResearchGate, which is sort of rogue in that publishers or authors can deposit a version of the article that may not be a true open access version, and then Sci-Hub where content is pirated. We looked at four different access types across these search locations. There is gold open access, which we defined very broadly as any version that we could get to a free version on the publisher’s website. Green open access: We looked in institutional and subject repositories, as well as on author websites, discoverable through Google or Google Scholar. A rogue version is anything that we found on ResearchGate. We did not try to go into ResearchGate and determine which things were legitimate open access versus rogue, so we’re just saying if it is on ResearchGate, it is rogue. Pirated means anything on Sci-Hub. Again, on Sci-Hub, some of it is actually open access content. Some of it is content that should not be available.
We searched each article by title in Google Scholar and in Google. We just did a title search. We didn't do anything further than the title search. We counted the access types. We counted in Google and in Google Scholar whether it was available in gold or green or rogue. In many cases, Google Scholar turns up ResearchGate or Academia.edu results. We counted the number of title match results in each. We counted the number of results with available full text, so how many things could we find full text for when we were not on our campuses using our licensed content? We then searched each article title again in ResearchGate because sometimes ResearchGate turns up in Google Scholar. Sometimes it doesn’t, so we searched directly in ResearchGate as well. We searched in Sci-Hub. And then we measured the title match versus the freely available full text results. So, we gathered a bunch of data, and now John is going to come up and talk about some of our results.

John McDonald: Thanks, Michael. This is the best part of the presentation, so, I’m the lucky guy that gets to give you guys all the results. For access type, again, Michael told you access type or, in other words, the source of the full text article, whether it was green, gold, rogue, or pirated, was our first set of results. As far as gold, green, and rogue, we had just a few simple research questions. Basically, how many are gold out of our article sample? How many are green, and then where are they green? Are they green in institutional repositories, subject repositories, or on author websites? And then how many are in the rogue and the pirate systems? For rogue systems, we did ResearchGate and Academia.edu, and for pirated, it was Sci-Hub. And a note about Academia.edu: You can’t search it directly, so we only got results through Google results, so you’ll see the results in one of the next slides.

So, here is the verdict. Out of our sample articles available in gold OA, we found that a total of 80 out of 300 articles were available gold OA on the publisher’s website. That’s 26% of the sample, and across the disciplines, it ranged from a nonsurprising 23% in Arts and Humanities up to 32% in the Life Sciences (see Figure 3).

John McDonald: Thanks, Michael. This is the best part of the presentation, so, I’m the lucky guy that gets to give you guys all the results. For access type, again, Michael told you access type or, in other words, the source of the full text article, whether it was green, gold, rogue, or pirated, was our first set of results. As far as gold, green, and rogue, we had just a few simple research questions. Basically, how many are gold out of our article sample? How many are green, and then where are they green? Are they green in institutional repositories, subject repositories, or on author websites? And then how many are in the rogue and the pirate systems? For rogue systems, we did ResearchGate and Academia.edu, and for pirated, it was Sci-Hub. And a note about Academia.edu: You can’t search it directly, so we only got results through Google results, so you’ll see the results in one of the next slides.

So, here is the verdict. Out of our sample articles available in gold OA, we found that a total of 80 out of 300 articles were available gold OA on the publisher’s website. That’s 26% of the sample, and across the disciplines, it ranged from a nonsurprising 23% in Arts and Humanities up to 32% in the Life Sciences (see Figure 3).
were not as discoverable as we expected. Subject repositories were a little bit better but still not great at 14% overall, and not surprisingly probably to all of the librarians in the room, the author websites self-archived were not very discoverable at all. We only found 10 articles out of our sample in total (see Figure 4).

As far as our rogue systems, ResearchGate and Academia.edu, we found that 30% of the total sample was available via ResearchGate, and the Arts and Humanities are not very accessible in ResearchGate as open access versions, but the Social Sciences ended up with 36% and Life Sciences 44%, so probably what everybody would expect. As far as Academia.edu, again, I didn’t put a percentage on the table here because we weren’t accessing Academia.edu directly, so there could be additional items in there that are open access, but this is what we got from our Google and Google Scholar results. Overall, the total for both of these rogue systems together were 111 articles, so 37% (see Figure 5).

And the grand total for all open access sources ended up being 166 of the 300 articles; we could find at least one version of an open access article. Arts and Humanities was just below 50%, Social Sciences very high at 60%, and then Life Sciences at 57%. And these results match the earlier research results that have been published in the literature that write about 50%, 50 to 60% of recently published literature is available in an open access form (see Figure 6).

<table>
<thead>
<tr>
<th>Discipline</th>
<th>Institutional Repository</th>
<th>Subject Repository</th>
<th>Author Website (Self-Archived)</th>
<th>Total Articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arts &amp; Humanities</td>
<td>6</td>
<td>4</td>
<td>5</td>
<td>13</td>
</tr>
<tr>
<td>Social Sciences</td>
<td>14</td>
<td>10</td>
<td>3</td>
<td>19</td>
</tr>
<tr>
<td>Life Sciences</td>
<td>7</td>
<td>27</td>
<td>2</td>
<td>27</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>27 (9%)</strong></td>
<td><strong>41 (14%)</strong></td>
<td><strong>10 (3%)</strong></td>
<td><strong>59 (20%)</strong></td>
</tr>
</tbody>
</table>

Figure 4. Articles available via green OA.

<table>
<thead>
<tr>
<th>Discipline</th>
<th>ResearchGate</th>
<th>academia.edu</th>
<th>Total Rogue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arts &amp; Humanities</td>
<td>11</td>
<td>20</td>
<td>26</td>
</tr>
<tr>
<td>Social Sciences</td>
<td>36</td>
<td>9</td>
<td>40</td>
</tr>
<tr>
<td>Life Sciences</td>
<td>44</td>
<td>5</td>
<td>45</td>
</tr>
<tr>
<td><strong>ALL</strong></td>
<td><strong>91 (30%)</strong></td>
<td><strong>34</strong></td>
<td><strong>111 (37%)</strong></td>
</tr>
</tbody>
</table>

Figure 5. Articles available in rogue systems.
To contrast that with Sci-Hub, we searched all of the articles in Sci-Hub, and we came up with an astounding 87% of the articles were available in Sci-Hub and equally across all the disciplines. We found 86 of our article in Arts and Humanities were available in Sci-Hub, and 87 in Social Sciences, and 87 in the Life Sciences (see Figure 7).

Looking at this availability then as a bar chart (see Figure 8), on one slide you can see then that gold open access via publisher websites, we ended up with 80 of the articles total. Green open access in all locations was not as available as gold open access, but ResearchGate in the blue bar—ResearchGate and Academia.edu actually performed pretty well with 111 of the articles. Overall, the Arts and Humanities are not well served by ResearchGate and Academia.edu but pretty comparable in gold open access at least. The Life Sciences have higher percentages, as most people would expect, but the Social Sciences performed pretty well, especially in ResearchGate.
Then we added the black bars here to show those compared to Sci-Hub, and you can see that 260 total articles, the 86, 87, and 87 across the three broad disciplines.

Now, those were the total articles. We also wanted to look at the additive availability by the article source so, for example, gold open access we found 80 articles by gold open access to, if as publishers and librarians, we feel like that is the most legitimate variety of open access that there is, 80 of the articles were available gold open access. And then if you start to look at things that were green open access but not gold open access, so how many additional articles were available in an open access version that weren’t available gold, but they were available green? We found that additional 24 articles, so then we are up to 104 out of our 300 article sample.

Moving forward, we looked at what was available in our rogue systems that wasn’t otherwise available in gold or green, and we found an additional 59 articles. Then if you go—we found 59 in the rogue system, and then if you add in Sci-Hub to complete your journal article searching, then you found an additional 115. Overall, all versions of freely accessible journals ended up over 90%, so our users could relatively easily discover about 90% of the articles in our sample. And you will see that even though Arts and Humanities is not as well represented in gold, green, and the rogue systems, Sci-Hub makes up for it with great coverage of the Arts and Humanities as well. Hey, if you’re going to steal articles, you might as well do it from the Arts and Humanities journals, too, right?

We also wanted to look at, as Michael told you earlier, we were looking at search location. So, generally looking at how users, scholars, mostly faculty and students, are actually finding this content. We wanted to look at Google Scholar, Google, ResearchGate, and Sci-Hub as the search location for all of the articles. And a little note about methodology, we did start off with Google Scholar, making a broad assumption that most academics know Google Scholar and may start with Google Scholar. Some institutions even use Google Scholar as their discovery system. We started with Google Scholar, and we were looking at search results for our articles, and we looked at the “All Versions” button below every article. They collate all the versions that they think they found, that Google Scholar thinks are the same article, and they put them together. So we found the results, and then we expanded to look at all 10 versions, and we also noted the PDF view. Google Scholar is promoting access to freely available articles and legitimate open access by directly linking to PDFs that they can find. You will find that on the right-hand side of...
search results. And we found that the “All Versions” for most articles out of Google Scholar, the overall average was 3.74. So Google Scholar is finding three to four versions of every single article. Unsurprisingly, the Arts and Humanities is not as well represented with only 2.5, and the Life Sciences much better represented with five. And then we found that Google Scholar will provide you access as a search location to over 40% of the journal articles in our sample, so you can get to it open access from Google Scholar for 122 of our articles.

And then when we then progressed looking at doing the same searches in Google, we found that it was the exact same number of articles that you can find through Google, 122 of our 300, and they were not always the same articles. So, the 122 we found in Google Scholar were a different set than you could find in Google. So, that’s why users should actually search through both of them. Fewer number of title matches in Google; they don’t collate the matches, but when you do article level searching here, you will see multiple versions come up and in the Arts and Humanities. It was just below 3, Social Sciences right at 3, and life sciences at above 3.5.

Looking at these results, availability by search location in one chart (see Figure 9), again Google Scholar is the blue bars, and we found 122 of our articles overall, Google with also 122, ResearchGate was 91 articles we found, and we put Sci-Hub on here also to underscore the total volume that you can get through Sci-Hub is 260. Google Scholar and Google operate almost equally in the discoverability of this content, and ResearchGate functions really well for the Social Sciences. You will see there were 36 articles found through ResearchGate in the Social Sciences as compared to 39 in Google Scholar and 40 in Google, and in the Life Sciences, it is even closer. ResearchGate does not have great coverage in content in the Arts and Humanities right now.

Looking at the additive availability by search location again, and we’ve got two different versions we’re going to go through here. Google Scholar provided access to 122 of the articles. If you then move on to Google and limit out the ones you’ve already found, you find an additional 32 through Google, so 154 now of our sample, so just over 50% of the content was available by just searching Google Scholar and Google. Then if you move on to ResearchGate, you’ll find eight additional articles that you didn’t discover before, and then again Sci-Hub, you will end up finding basically the rest of the sample. So, up to over 90% of the total articles were available if you use all four of these methods.

Figure 9. Availability by search location.
Moving backwards, though, we wanted to look at it if authors, faculty that are on ResearchGate and use ResearchGate as a discovery mechanism, if they actually started from ResearchGate, what is the additive availability by using ResearchGate first, and then moving on to others. You would find 91 of the articles via ResearchGate, and then when you go to Google Scholar, you find an additional 54, an additional 18 from Google, and then basically the rest from Sci-Hub. I'll turn it over to Jason to conclude.

**Jason Price**: John says his was the best part, but I think this is the best part, although it is also a sensitive subject in some ways. I want to be a little provocative and lead into some discussion, so we're looking forward to that.

So, in conclusion, it is hard to follow the rules. If you stick with the open access versions of articles, you are limited to somewhere in between 20 and 40% depending on the source of that version, and in fact, I guess 20 to 25% on the classic rights-appropriate open access. If you go into the rogue open access, which is potentially much less rights-appropriate, but you increase the number, but if you want to go just one place and get the most possible freely available articles, as a researcher who doesn’t think that it is wrong to download pirated articles, you’re going to go to Sci-Hub. Starting with Google Scholar and supplementing that by Google is slightly a better strategy than starting with ResearchGate, but you can kind of move both ways, and even though you see 30, 20, 40%, you can get up to a higher number if you use one and then progress on to the next, if need be. Again, starting with Sci-Hub and bypassing the legitimate search options entirely gives the quickest and best results. And an obvious conclusion from that is that libraries and publishers should be concerned if our users decide to go here instead of using the contents we are licensing, that is a huge problem, one that we need to recognize and not ignore.

Before I go on to some of the potential applications of this, I want to talk about one next step that we haven’t taken that we think is really important, and that is to examine both OA discoverability and availability in library discovery systems. So, this graph (see Figure 10) looks at the four most popular discovery systems, and the blue bars are articles that aren’t available in open access, and the gray bars are those that are. So, the question is if you just drop that title in that discovery system, is it going to be indexed? This is not necessarily a test of OA versus non-OA, but that’s the intent: Is OA content less well indexed in library discovery systems? That is a relatively important question. We didn’t see strong trends toward that, but we did have this discoverability side, although there could be something underlying this. More importantly, potentially, is how effective are library linking tools at providing the full text access to open access articles? So, if you find it in your discovery system but you don’t have licensed access to it, how commonly do our systems lead to that full text? We expect the answer to be not nearly as commonly as they are actually available out there on the web, but we would like to actually design a study to look at that in a little more detail. And I think we, I work for a library consortium of very small libraries, many of whom do not have site license access to a lot of this content, and I think doing this work with them, examining that some more, might open up some possibilities.

The theme of the conference: “Roll With the Times or the Times Will Roll Over You.” This theme and our presentation I think really fit well this year. The times, led by faculty who are sharing articles in ways which may or may not be rights-appropriate and who feel like it is fine to download pirated papers and are going there, that’s the times are pushing forward, and we need to not ignore these things. We need to recognize them and think about how we can react and respond appropriately.

I have three puns for you, and I’m going to give an example of each. The first is “Collar Google Scholar?” The second is “Emulate ResearchGate?” And the third thing to do in response this is: “Don’t ignore that there is a Sci-Hub Pirate Club out there.”
“Collar Google Scholar,” what do I mean? Maybe we should be linking to Google Scholar results from our open URL resolvers in order to leverage more open access full text. That is a possibility and/or drawing Scholar open access text links into the results menu when they are available. Google Scholar is actually doing this. They have created a plug-in which allows you to highlight text, hit a button, and then pull up this sub-window on the right-hand side, and that green button is an open access button. If you are a researcher and you've added this plug-in, which they are now advertising underneath their search results, they’re making it obvious to faculty that this exists, and they are leveraging and making these open access links much more visible and likely to be used by researchers. They are even going to the point where on a publisher page where it says, “Purchase this Article,” they are pointing out that potentially, even without selecting any text, if you hit that Google Scholar extension button, it shows you the open access version of that article instead of the one that you might pay for as a researcher. So, we need to recognize that Google Scholar is leveraging these links, and we need to find ways to leverage these links. I just learned today, actually, that Elsevier has created an article-level knowledge base that indicates which of the articles are freely available and which are not. So, think hybrid journals: You can’t use the title and figure out whether it is open access or not because there’s both kinds in there. They have an API which is freely available to folks to potentially, if you have a DOI, you check it and it will say, “Yes, this is open access,” or “No, it’s not.” We could put an article level link in our results pages for folks who don’t subscribe to those journals but can get access to them. These are the kinds of things that I think we need to be doing to be keeping up with the times.

Second example: “Emulate ResearchGate.” So, this is something a library is already doing: Include metadata for all faculty publications in institutional repositories, even if the OA copy is not available and even potentially if it never will be, and allow users to request a copy through the institutional repository listing. So, the text on this is small, but you’ll get the idea. This is the University of Liege (see Figure 11). This is their institutional repository. They just have an abstract. The bottom of the page on the left-hand side shows that there is restricted access to this article, but there is a PDF in there. On the right-hand side, they have a button to request a copy. Does that sound familiar? That is what you do in ResearchGate. Here is a library doing that. When you hit that button, it tells you if you are from the University login. If you are not, here are the rules, but you can ask the author of this article for a copy. That’s what they are doing with their institutional repository. I think this is emulating the fact that ResearchGate actually covers—it has listed—nearly 100% of the articles we looked at. I think that is important otherwise our institutional repositories really don’t cover an extensive portion of our faculty publications.
The third example is less of an example and just something that when we found this as part of doing this research, I was floored. Here is a big thing that is going on that I think we should know about and recognize is happening. Remember 88% of researchers did not think it is wrong to download pirated papers, and 87% of the papers are pirated and available through Sci-Hub. 87%, right? There is a plug-in that if you go to Sci-Hub’s site, if you look for an article and it’s not found, it gives you a link to install this Google Chrome extension. Now, that said, it is a developer mode. It’s a little funky kind of thing, but because Google has not endorsed supporting Sci-Hub, they are not adding this extension into their publicly available content. What you’ll notice if you look closely is that down, you probably can’t see the URL, but it points to the article from a Google Scholar interface in Sci-Hub. When you click that title there in a Google Scholar interface, you go directly to a Sci-Hub pirated version of that article instead of going to where you normally. If you had the Google Scholar without the plug-in, you would go to your campus’s licensed access if you’re on campus. This makes it extremely convenient to access 87% of the articles published in 2015 across the disciplines. That is scary to me but also something that I think we can’t ignore and need to address. So, with that, I’ll open it up for questions, comments, thoughts.

Figure 11. University of Liege institutional repository.
Building the Knowledge School

R. David Lankes, Director, School of Library & Information Science, University of South Carolina

R. David Lankes: Thank you. Alright. This, by the way (referring to slide that says “In Search of Geeks with Social Skills”), is our marketing campaign for our undergraduate program. If you know potentially people that are not geeks but still have social skills, we have waivers available to them. There is a sort of a story that goes with this if you will excuse me introducing it. This came from one of our alumni. We’ve been talking about what is the knowledge school and where we’re going in our school, and we’ve been talking with our alumni, and going through what should be, and where it is, etc. And the fellow said, “Well, what you’re really looking for is geeks with social skills.” And I said “Yes, that’s it! That’s amazing! Absolutely!” And as you’ll see in a moment, we have a beautiful building on the University of South Carolina campus. It is sort of not technically on the horseshoe, but it’s close enough that we pretend it’s on the horseshoe, and it’s got these massive columns, 20-feet high columns, and I thought, “We’re going to put this on the columns!” So, I showed it to a few people, and they said, “Oh, that’d be great,” and I showed my Provost, and he said, “Oh, that’s really kind of funny.” Then they said, “But, could you just run it by the engineers?” I said “Sure,” so the Associate Dean for Engineering said “Oh, fine, marketing. We don’t care. It’s yours.” And the folks in Computer Science said, “Eeeeh, it’s fine.” But, the Dean said, “We will go to the Provost and will have a debate about the word ‘geek.’” I’m like, “Seriously?” And so, it is now become a benchmark that if you find this an attractive slogan, you should come into our school, and if you find this offensive, you can go into engineering. I think they were more used to being called this in a derogatory way. That’s not what we’re about.

Hi, my name is David Lankes, and I would like to welcome you to my newly adopted state. I’ve been a citizen of South Carolina now for four months. I moved from Syracuse, New York, to the University of South Carolina where I am the Director of the School for Library and Information Science. And if you’re wondering why I made that move, perhaps you could spend a little bit more time outside. Though people ask, “Are you adjusting? Is it what you expected?” I did not expect to come in August and have 10 days of straight 100-degree-plus weather with the necessary humidity to go forward. I didn’t expect a tropical storm. I didn’t expect a hurricane. I didn’t expect my Dean to step down at the same year, but other than that, yeah, it’s going really well. But, we’ve been talking, and I realized as I was preparing this talk and the organizers were kind enough to give me a slot, that this is my third speaking engagement at Charleston, and so I realize that I’ve done a trilogy now, and I’ve realized sort of on a personal journey that each of these presentations has come at a very instrumental time in my thinking and in my career in such, and this one is no different.

In 2006, I came and talked about massive scale librarianship, and the idea was, as we heard this morning over and over and over again, it turns out we as human beings are really good at producing information and really lousy at capturing it all. And so, I like to think that I started that conversation in 2006 when it was probably 4,000 years ago when someone said, “Could you give it up with the scrolls? We’ve got enough!” What I realized at that time, that was a sort of realization as we were thinking about librarianship, and we were thinking about collections and were thinking about materials and acquisitions and roles with publishers, about really this notion of a hybrid collection, that we had to acknowledge and understand that ultimately our collections weren’t what we licensed, weren’t what we owned and purchased, but in essence they were software and all the things once again we’ve heard this morning. And that’s evolved to really the collections and what we are preparing librarians and information professionals to deal with is that the true collection of any library is the community itself. The books, the materials, the databases, the emulation software, the archives, all of these are tools, and tools to help develop that community and move that community forward, but that knowledge, and this is one thing that I did my best, I sat really, literally in the back of the room today and tried not to jump up every time, although the people sitting next to me did notice a few of these (pretending to twitch) every time they talked about knowledge as something that you could put in a binder and put on a shelf. Because if you think about it for a moment, those are materials, those are interesting things, but knowledge is uniquely what is in our head.
If I give you a book and it’s written in Chinese, and you don’t speak Chinese, can we truly call it knowledge? We can call it “capability of knowledge.” You can call it sort of “latent knowledge.” It’s waiting for engagement, but it isn’t until we take it up within the community and try to apply it to our context and our situation that it becomes knowledge. And this became very, very clear to me recently, and for those of you who have just come from the plenary session, the slain librarian is my alumni. She came from my program. We have recently—she hosted co-ops for people up the street, and what I have heard from people that have worked for her, who have learned from her, who she was a mentor to, they have said the first thing she would do is she would put us in a car and drive us around the community. The first thing she would do is she would go out to where the communities are, where they couldn’t necessarily get to the library, and her message was always, “This is your library. This is the community.” That knowledge is what that community needs to move ahead, what that community needs to advance. That knowledge is not a cold thing is. It is not something that sits on a shelf or repository or an archive. It is passion. It is light. It is understanding, and it is an intensely human thing. The issues of how we capture data, how we capture materials: Vital, important, absolutely. Core to what we do, but let us never mistake that what we are collecting them for is not for the sake of collection, and while I love the concept that we are in the business of eternity, we have an obligation to those in the present to figure out how to help them improve their life. So, that was 2006, and we called it “Participatory Librarianship.” And then in 2009, they brought me back, and I talked about new librarianship and in it, this was, I looked it up, this was the first time that I sort of publicly put out this concept that the mission of librarians is to improve society through facilitating knowledge creation in their communities. That has since become something that has turned into the Atlas of New Librarianship and additional books, and actually it was with my same moderator who then looked at me and asked the first question after it, and he said, “As a publisher, that is my mission.” And I’ve had teachers say, “That is my mission.” And I’ve had lots of people, academics say, “That is my mission.” Google could say, “That is my mission,” and I have to say when he asked that question, I gulped a little and then thought a lot later, and I said, you know, that’s good. Having an open mission like that means we have allies, and it means that we have people that we can work together. So, that began a whole different thinking, and now in 2016, I am back because really this is the next step in this evolution.

The first step is it’s more than just stuff. It’s communities. Its people. Its knowledge. It’s human. It’s understanding these things. The second was our role as librarians is we must facilitate this, be part of this, and help it push forward. We must work with publishers and data and scientists and our communities as a collection, and we must figure out how to help communities make better decisions, how to learn. And now, I’m here to say it is my turn to start talking about how we push that forward in a very specific way. We all have things that we can do, and one of the things that I am now very essentially concerned about, and my faculty are very essentially concerned about, and my staff are essentially concerned is what role does a library school or a library and information school play in this ecosystem? What does it look like to prepare librarians today? In times of radical change and in times of the mutating library, how do we prepare people for this? How do we come out, and how do we deal with the fact that we are now generating people who walk around and say, “I’m an information professional,” and everyone looks and goes, “... and an information professional is ...?” That geeks with social skills came because we have a Bachelor of Science and information science. Can I just tell you how excited 18-year-olds get by saying, “I’m going to be an information scientist.” Can I tell you how excited their parents are when they ask, “And what’s that job title look like?” And we’re like, “Eeehhh ...” But they can be librarians, and can I tell you the rare unicorn who is an 18-year-old that says, “I’ve always known I want to be a librarian, and I’m going to start now even though when I get the bachelor’s degree it doesn’t count.” These are the things that we are wrestling with. So, what I’d like to do is I’d like to talk about how we see and envision building this knowledge school and really as a way to begin a conversation.

So, with that, I want to give you a little history because one could say, “We already know how to do this, Dave. Come on, we have library and information schools. We have iSchools!” Many of you are probably graduates of or related to or, as case may be, you’re graduates of a library school that is now an information school, and do you know why and are you happy with that? Did they change the name? And
all these things. So, we could simply say “Let’s see where we bend,” because clearly this is the blueprint. We can see in 1988 we talked about the “Gang of Three,” and the “Gang of Three” was Syracuse, Pittsburgh, and Drexel. We all hung out together at ACES Conferences, and we all gave each other high fives at how advanced we were, and it was really cool and we were better than the other kids, and frankly, it was all an evil scheme to overthrow the University of North Carolina at Chapel Hill and Illinois because those bastards owned the rankings, and we to this day are waiting for their alumni to die out. So, good. Now, see, we are loosening up?

This turned into, by the 1990s, we had two years to work on it, and we added a person. It was Rutgers that said, “No, no, we too! We, too! We can do communications!” And that’s cool. So, Rutgers began playing, and we were fine because we had places to stay when we visited New York City. Then by 2001, it took 11 years, and we added one more person, we were working hard, and so now Washington. Now, why Washington? My adviser, Mike Eisenberg, who I love dearly, quit, took the job at University of Washington, and he used to call it, and I’m going to remind him of this every time I see it, that University of Washington was the “Syracuse West” program. He doesn’t say that anymore, but we began to expand and think about this. Right? We began looking at schools moving from schools of librarianship to schools of library, and information science to iSchools to schools of information, etc. Then the “Gang of 10.” I don’t know why it was the same year, but apparently they were really quick this time, we brought in Michigan and Florida State and UNC. Now we have the “iSchool Caucus” that started in 2008 where we have a ton of schools that identify themselves as iSchools. So, this is a sort of chronology, and we can say, “Well, this is easy, Dave. You’re South Carolina. You write them a check. You become the iSchool caucus, and you’re an iSchool.” And we know what that means. That means go get more research funding, build a big undergraduate program, do lots of flashy things. We’ve already put a big TV in our hallway; it’s exciting.

We can also look at this evolution from a structural standpoint, as a topical standpoint. This is my take, a very unscientific take, of the evolution of where we began as library science. If you look at sort of the iSchool movement, it began as library schools that begin thinking differently. Back in the mid-70s, Robert Taylor, when he became the Dean of Syracuse University, renamed it from the School of Librarianship to the School of Information Studies. By the way, why “studies” and not “science”? Because everyone admitted we had no idea what it meant, and science sounded too pretentious. And that’s why, by the way, I have a PhD in Information Transfer. I told this to one of my college buddies who instantly said, “So you’re getting a PhD in being a bike messenger?” I said, “Yeah, pretty much.” And information library science, we sort of knew what it was, it was cataloging and materials, and, yes, we could talk about collection size, and we could talk about different schemas for organization. And that really then grew into this notion of Information Studies, which was happening outside of libraries, and we saw more and more organizations looking at information as a strategic asset. We saw it in the corporate sector. We saw an explosion in the government sector around information resource management, if those of you lived back in those times. We saw the advent of information CIOs, and so that happened. We saw that library science grabbed technology very early on and very aggressively. An outgrowth of that became information retrieval. It brought computer scientists and librarians together. We had lots of data. They had lots of time. It worked out. Luckily, we found people with lots of funding. So, information retrieval became core, and what we begin to see as information retrieval has been advanced to the larger concept of information technology computing, human-computer interaction. Information studies became a strong emphasis on management and increased in their own communications. We see this sort of broadening of the conversation of the topics.

Now, what’s also interesting is that I’m going to give you one more set of evolutions that’s on this rough timeline. Really, when we were talking about this in library science land in early times, the focus was on professional preparation. What do librarians need to know? What is an information professional? What are their core skills? That led to the Golden Age. The Internet, we have a moment where we say, “Boy, that’s really changed our lives now!” But you have never seen more red meat put in front of hungry dogs than when the Internet hit the information schools. Suddenly, it was all about experimentation. It was all about what can we do with computers? What can we do with technology? We can totally change health care. We can totally change libraries. We can totally change this. And everything was put in front of digital. Right? We had this acronym of
virtual and digital, and it became virtual libraries, virtual reference. I was there. Don’t blame me. Well, blame me. Digital libraries, etc., and everything was new and cool, and it was an amazing sense of expectation. And it grew very much in this “Gang of Five” era to a sense of social change. There was a real sense that we could use technology, that this new larger concept of information could talk about societal change. And we’re going to come back to that, but at this time, it’s worth noting that it was almost always from the perspective of sort of technological determinism, that is society would change of course in a positive way because applying technology always has a positive effect.

How many of you have ever bitten your tongue when you’ve heard that person say, “I am going to solve world hunger. I’ve written an app.” And you’re like, unless they could eat the damn phone, you have not solved world hunger, but that is really where we were. I was, I am not making this up, I was the coordinator of a virtual reality laboratory in 1989. I can tell you that I could take the press releases and the garbage I wrote then, and put it out now and just put “Samsung” on the top of it, and it would be the same stuff: This concept that things were going to fundamentally change, and it was always going to be for the better. What we see now, particularly among the iSchools, is a much different conversation and its institutionalization. Part of it is a respect thing. How do we get people to respect us? Are we a field unto ourselves? Do we have to look at the historians and say, “Look, we have theory, too?” But a lot of it is how do we build in support? How many faculty do we need? What is the faculty-to-student ratio that makes sense? How do we staff that? How do we support it? So, there are a lot of discussions about what is an iSchool and a lot of discussions about information, but I would argue that there is also a lot of caution that I have perceived, primarily from the library community, but also on my own. And that is a growing “L” versus “I” breach. That this also goes in cycles, and it began with the “L” word. “Oh, no; you are no longer a School of Library Science.” By the way, I did every time I heard the word “user.” How many people think about that you serve your users well? How many people enjoy being used? Think about that. When you conceptualize, talk about, and relate people as users, what you are saying is you are putting yourself in a position where you are used, and they are using you. And that might be okay, but don’t ever talk to me about consumers. Because on the information side, what do we see? We see it is all about technology. We can do this with technology. We have these ubiquitous networks, right? We see that instead of values there is sort of a social science perspective that is that it is value-less. That we go to a community, and we understand what they are doing, and we have objective measures, and we bring qualitative methodologies to it. But there is an antiseptic nature to the approaches around social science sometimes that doesn’t come from anthropology and sociology, but comes very much from this concept of when we move from the humanities view of libraries to the social science perspective and technical perspective. We begin to address this concept of objectivity, and it became problematic. A focus instead of on service on products. What is the next app? What is the next solution? What is the next brand? And once again talking about users: User experience, user-based design, user, user, user. And I’m going to argue that what we need to do is put these back together. That neither of these are wrong, and, in fact, the librarians and the information professionals need each other because this should be a unified set of skills and understanding, not separated.
And so how do we put that together? And that’s what I want to begin talking about in this concept that we as South Carolina are calling the “Knowledge School.” And there are three components that I want to talk to you today about this Knowledge School concept. The first is how we move from a school to a school of thought. How we move, and how we must be focused on participation and impact. And how we need a unified mission but diverse in how it is implemented in our communities and the delivery method. So, let’s begin here by talking about moving from a school to a school of thought. What is it that holds the school? Think about for a moment your alum, your alma mater, whether it was South Carolina, whether it was Syracuse, Emporia, whatever it was. I don’t care. What is it that has made the community better by them being there? What held them together? Why did those professors with those disciplines sit together? Now, sometimes the answer is because they were thrown there. Sometimes it just evolved that way. Sometimes it was opportunistic. The question becomes why are we together now? One of my faculty members, fabulous faculty member, Jennifer Arns, is in the audience. Why is Jennifer sitting next to me? And why does Jennifer, whose great work in the area of public policy and impact of public libraries, why did she sit across the hallway from someone who’s doing data mining and looking at Twitter feeds? And why does she sit upstairs from someone who is working in school libraries and basic literacy? And why are they sitting next to someone who is working on information within religious communities? Why are we together? And the short answer can’t be because we are preparing librarians. Because we know that our alums, yes, some become librarians, but they also work for Google, and they also work for Springer, and they also work in lots of different places. And our undergraduates, some of them don’t work in libraries either. Why are we together? And that must come from the school of thought. What I’m going to argue is, rather than saying it’s because that’s where our vita was, or that was our history, we must do it because we are all trying to solve a common problem. So, as we begin to think of it, what is the common problem that holds us all together as library and information science?

I’m not the only one to ask this. We brought together a group of people: Andrew Dillon, Sam Hastings, Tula Giannini from Pratt, Anne Craig, Annie Norman, a bunch of people. We said, “What is the grand challenge of library and information science? What is it that we are trying to solve?” Bill Arms, who was at Cornell, had a great thought experiment. He said, “You are the new president of your University, and you’ve asked every school to send you their star, that is someone who is highly recognized and is highly influential and is recognized by other people.” So, the math department sends someone who got the whatever prize, and the business sends the guy who got the Nobel Prize in economics, right? And all of these awards show up. What does the school of library and information science’s star look like? What problem did they solve that was so fundamental that they made an impression, and that impact is also recognized by all the other people sitting around? What is it that we do that physicists go, “Thank God you’re here?” What is it that we do that philosophers and historians say, “That was a really good job you did.” And so that is the notion of a grand challenge. Now what’s interesting about a grand challenge is it’s not about us. A lot of the dialogue that we see around the future of library, and that we see in the iSchool movements, and we see and hear all the time are why are we sitting together? And it’s very introverted. What we need to talk about is: What is the problem in a community that we can provide the solution for? What is it that we have solved that our communities need? And so this group conceptualized, and they said what we live in a knowledge society. By the way, this gets really practical at the end, I promise. Stick with me. The knowledge society is the fact that we don’t live in an information society. We’ve grown past that. It’s not a matter of access to bits and data and stuff. It’s access to intelligence and good decision-making, right? It’s not about how much we can push at people but how much they can make sense of the things that they need, right? Our problem in this presidential election is not that we have too little to read. It’s the fact that there is so much to read that you can pick what you want and that when you read that there is a lot of heat and very little light.

So, how do we bring that together? And understanding that in this knowledge society, one of the fundamental differences between this view of the current world and the world of 10 years ago, frankly, the world that we heard in our opening plenaries when we talked about what the users want. Do we believe that there is a single unified concept of “user” that we can all serve? And by the way, back to “used.” Do we want to talk about the people that we can give things to, or do we want to talk about partners and...
communities and members and such? And understand
that the diversity that we work at means that, yes, we
should look at unified economies of scale, but
ultimately our abilities to match very local needs. And
so what is big and different in this world is that we
now have to accept that there are multiple ways of
coming to a solution or the truth. How do we function
in a world that truly believes there are multiple ways
of arriving at the truth and that those truths can be
very different? How do we function? Is our job to fix it?
Is our job to ignore it? How do we deal with this? How
do we deal with the fact that we can look at data and
walk away and say, “That is bogus. It’s rigged. It’s
fixed.” How do we deal with that? In this knowledge
society, we have a functioning knowledge economy.
That is, how do we then work? How do we develop
and distribute our resources? The input to this
economy is through innovation, progress, and a
workforce; the people who we are preparing to go out
and work. The 3-year-olds, the 10-year-olds, the 60-
year-olds that are constantly having to figure out—
how do we divide these resources and the knowledge
economy takes these resources and decides how we
distribute it? How we define the community, and how
we are constantly learning in this environment? That is
the society that we are functioning in. The grand
challenge is how, based on that economy, it functions
on an infrastructure just like we talk about market
economy and transportation economy, right? Getting
goods from one place to another needs our roads, our
infrastructure, getting ideas and thoughts and
understanding and learning from one place to another
needs an infrastructure. And that infrastructure
consists of technology, and that’s the obvious part.
That’s the pretty part, but it very much consists of
sources. That’s where a lot of the focus of this
conference and folks are in, the things we use to make
decisions, the materials, the resources, the reports,
the data. It also involves permissions, that is, who can
get to those resources, open access, copyright, right?
Fee for pay, all of these things. And finally, the
people—the people that are making the decisions.

When we look at this knowledge infrastructure, we
have questions. We note that this current
knowledge infrastructure is currently uncoordinated
and conflicted. That is, there is a lot of people doing
a lot of stuff, but no one is talking to each other. I
very much appreciate the phrase “information policy
wars.” That is not too much of a metaphor. We have
conflicts between people who want to make money
and people who do access. We at this conference
represent the amazing conflict. We have people who
are trying to sell us stuff, while other people are
trying to give it away, and we are all trying to get
together and figure out we can still be happy about
that. So, a little conflict is always going to be there,
but uncoordinated makes it problematic. If you have
ever traveled overseas and you’ve ever tried to get a
SIM card, you know how problematic it can be. I was
over in the U.K. I was in London, and I went into a
Three store. I said I would like a SIM card. And five
minutes later, five minutes later, I walked out for 30
pounds with unlimited data and unlimited
international calling, and it worked in my phone, and
I could walk out. If I walked into a Verizon or AT&T
store, three hours later I would be pissed off,
thought that I had been ripped off, have a phone
that doesn’t work, and by the way the SIM will come
next week and, etc. So, conflict, not always good.
This market economy thing needs some thinking. We
know that there are challenges to conceptualize and
form this infrastructure. We heard that today. We
talk about knowledge as if it were a thing, and we
talk about that thing as if it’s a book. I mean, call it
whatever you want, journal article, whatever. We
think that it’s a well-heeled, well-understood, well-
synthesized piece in front of us, and that’s not what
it is. For example, if you drove here today, there’s an
excellent chance you generated about a gig of data
in some computer somewhere. You generated it
because you used a GPS system. Maybe you listened
to music online, or you had an intelligence
management system telling you where. Is that part
of the infrastructure? And you sit there and say,
“Well, does that matter?” And I say, “Well, if you’re
in the information infrastructure business, are you in
the road business?” Remember the days when we
were arguing whether the Internet was part of our
collection? Now are going to talk about is the
highway part of my collection? And you sit there and
go, “Absolutely not!” But talk to a transportation
librarian. Number one use of tolled data, our FID
tolled data at New York State, is tolling, duh.
Number two: Divorce lawyers. If someone walks into
your library and says, “I need to find out where this
person was last Thursday.” You actually might need
to be going and querying not a database but a piece
of asphalt. How do we prepare people for that skill?

This will always be a marketplace of public and
private. We need to talk about how we move from
the consumption-production dichotomy to
participation. If we are constantly preparing
librarians and information professionals to be used
and users, consumers and producers, we are setting
up a generation of dependent and independent people in society. If what we’re talking about is, we take materials to underserved populations, poor kids in rural South Carolina, and we give them books, but we don’t inspire them to write their own, we are creating and furthering dependence on a system, not liberating and challenging the norm. Because you have writers, and you have readers. A famous illustrator said I used to go into kindergarten and ask, “How many people of you are artists?” And they all raised their hands. “How many people are writers?” They all raised their hands. He says I go into a third-grade classroom and ask the same question. No hands go up because they’ve been trained that they read. They have been trained that they watch, not that they produce; not that they make, but that they consume. We need to take on an infrastructure that allows not simply quick, fast, and interesting access to other people’s stuff. If we constantly argue about how do we build an open access model that simply allows people to consume things for free, and we don’t realize it is the “consume” part that is much more problematic than the “free” part, or at least as, we miss the boat. And many of us come from academia where we sort of assume people are producers. That is what our faculty are like all the time. But look at the assignments we give to our undergraduates and our graduate students where they’re showing up, and they’re taking and copying and pasting the abstract, hoping to God that they don’t put into turnitin.com and seeing they got the “A.” That’s not what we should be supporting.

So, our ultimate grand challenge that we need to solve is how do we coordinate a knowledge infrastructure to speed learning and improve the decision-making of our communities, right? That’s what we want to do. That person who shows up to the president’s office, their goal is not to say, “Boy, I figured out BIBFRAME last night.” Their goal is to say I helped your university, by the way, be 28% more productive, increase the GDP by this amount, help our students become these things. Our goal, our result, our grand theme is talking about a society that has truly open access to ideas where they move, and they understand, and they learn. Where we can bring the Trump supporter and the Clinton supporter together, and we can say, “Look you may not agree, but let us at least agree on what we’re going to discuss and what we are going to accept.” In this world, what then becomes the research agenda of a knowledge school, of mar researchers, what do I look at? Why is Jennifer next to these folks? Lifelong learning. We need the people who are speeding decision-making. We need to understand how they learn. People don’t read. They learn through reading. People don’t talk. They learn through talking. They’re constantly learning and adjusting their environment. We have to be aware and understand how lifelong learning. Why are our school librarians here? Because they are instrumental in understanding how we teach information literacy at the higher ed level and the community level and the special level.

The science of facilitation. We used to think that we could automate everything, but now we know that people who can afford it get people to do it. How do we facilitate this? How do we bring that conversation together? How do we bring communities in a land of multiple ways of the truth to still be a community? And so we talk about the idea of public policy. We talk about the idea of access for people with disabilities. Community: How do we bring communities together to do resources and finally moving from consumers to participants? This is an agenda of a knowledge school, and, yes, they’re going to look very differently. Someone is going to work in a classroom on basic literacy, etc., but it fits together. So, we move to a school of thought that conceptualizes our job as to improve the knowledge infrastructure and libraries and in business and in government and in not-for-profits. And then we need to say why do we do it? Because we are focused on practice and impact. Two images (referring to a slide): The one on the left is Cocky’s Reading Express, and that is Cocky. I am now a proud Gamecock. Thank you very much. And Cocky’s Reading Express, that is our school mascot. What happened is Sam Hastings and Kim Jeffcoat and the faculty of the school got together and said, “We have a problem. In rural South Carolina, we have a literacy problem. We need to solve it.” And so what they did is they bought a bus. They got a bunch of books, and they got Cocky, and they brought athletes and football players and undergraduates from across the discipline, and they drive them to Union. They drive them to Calhoun County. They get off there, and people show up. They’re excited. They give them pizza, and they give them a new book. And that connects them to self-worth, and it begins to talk about literacy, and they demonstrate that reading is important. Even if you want to be a football player, you got to read. Because we know that if you are not reading at grade level by third
grade, your ability to succeed in high school and the percentage of people who are going to drop out of school goes way up. We got to start early. The other picture is Columbia. I’ve actually been on that corner. Just about a year and a half ago, our 1,000-year flood. What happened when FEMA showed up is, they showed up, and they showed up in public libraries because they were community centers. That is where they could begin talking. And our faculty went out and said, “Okay, how do we deal with people with disabilities? When you talk about evacuating people from the coast, were they accessible? How do we deal with health information?” That is, did you know that water is full of sewage, and this is what you need to worry about in terms of health information? I had a great story of someone in this past hurricane of someone who couldn’t leave their house, not because it was flooded, but because the yard around it was flooded, and it had water moccasins flying around in it! Dear God, I live here now! The point being it’s not just about doing good research and sitting back, documenting and publishing. It’s doing. If we know that literacy works this way, get out and help people be literate! If we know that disaster relief should work that way, get out there and help in disaster relief! During the evacuation from Hurricane Matthew, a number of my faculty and staff opened up their house to strangers during the evacuation. That I consider part of being part of my school. That is commendable and rewardable service, that it is part and integrated into the access of impact.

Finally, we must talk about a unified mission but in diverse ways of talking about it. So, we have a thought. We’re going to fix the information infrastructure. We’re going to facilitate it and help people learn. We’re going to do it by actually doing things. We’re going to study. We’re going to write, documenting and publishing. It’s doing. If we know that literacy works this way, get out and help people be literate! If we know that disaster relief should work that way, get out there and help in disaster relief! During the evacuation from Hurricane Matthew, a number of my faculty and staff opened up their house to strangers during the evacuation. That I consider part of being part of my school. That is commendable and rewardable service, that it is part and integrated into the access of impact.

When Bob Taylor in the 70s showed up at Syracuse, he said, “I know information is important and increasingly important outside of libraries, and I’m going to make a bargain with librarians. If you go with me, if you support us going to Library and Information Science, if you support a master’s program, if you support technology and learning about technology, it will benefit libraries.” And I think at this point we actually have some success in this. We have libraries that are better enabled to participate in technology. DPLA would not have existed had we not brought information technology as part of librarianship. We
have values. We have ethics. We have lots of librarians out there. We have them working in different industries, and we get the attention of people like Google. Google comes and recruits from our staff. We have doctoral students going. That promise is made. What is the next promise? The next promise that I want to make, that I want us to be a part of, to get feedback and discussion as alumni and aspirational alumni and whatever, the next promise is: If we go beyond informing and information, if we go beyond a static view of simply providing big pipes to free information, if we move beyond consumers, if we move to truly participation, knowledge, learning, social action, and social engagement, we will improve the status of librarians in libraries, outside of libraries, and by doing so, we will improve the society itself.

And so, what I ask of you is, while you might already have your degree, send me your poor, your in need of status, your 18-year-olds lost, but send me your ideas and thoughts. Let us truly figure out how we can use our social skills and our unique capabilities of librarians to improve the society that we are a part of and how we educate and prepare and marshal troops of librarians and information professionals to make that happen. Thank you very much.
The Evolution of E-Books

Mitchell Davis, Founder & CEO, BiblioLabs
David Durant, Collection Development Librarian, East Carolina University
James O’Donnell, University Librarian, Arizona State University

The following is a transcript of a live presentation from the 2016 Charleston Conference.

James O’Donnell: Thank you, Tony, and good afternoon. My view of e-Books resembles Gandhi’s view of Western civilization. If you ask me what I think of e-books, I’m likely to say, being only slightly provocative, “Sounds like a good idea, and I hope somebody invents something like that someday.” Okay, I know I’m being provocative. On other days, I might just easily like to claim that I published the first scholarly monograph ever distributed over the Internet, a book about Dante’s Epistle to Cangrande which we distributed by Gopher back in 1994 from the University of Michigan Press. We have been not inventing the e-book for a good long time, but the challenge of the things we now call “e-books” is that they aren’t books, and they are only moderately “e,” and they just plain don’t work very well. I could speak to the mass market product, the ePub form, the Kindle form, which more resembles the papyrus roll of antiquity than anything else I know. As long as you’re willing to start on page 1 and scroll through to the end, they work pretty well. They are good for reading a murder mystery. Get to footnote number one, and you’re in trouble already. Try to look at map number three, and you are deep in trouble, and if you want to see the illustrations, the chances are they will be small and blurry and unintelligible if they are there at all. Those things can do something. They can’t do everything. At their very best, and I think the Kindle e-book is probably the e-book of the moment at its very best, they are considerably dysfunctional, and we could have a show of hands of how many of you believe that the Kindle format is the eBook of the moment at its very best, they are considerably dysfunctional, and we could have a show of hands of how many of you believe that the Kindle format is the eBook of the moment at its very best.
I am going to say concisely four things that I think we need in order to get to a point where we think the scholarly publishing system is designed to suit the print book in e-form, mainly therefore a legacy service. That is to say that it can take as many as possible of the features that generations and generations of compositors and printers have devised from making the printed book so functional and provide useful equivalents for them. I’ve got to be able to see the pictures. I’ve got to be able to read the legends on the map. I’ve got to be able to read the endnotes quickly and easily. I hate endnotes in printed books, but at least I have a thumb and forefinger and know how to flip back and forth. This technology is better than what any of our e-book vendors are providing us now for flipping back and forth to the endnotes at the back of one of their books. We, therefore, probably need another solution for the e-book going forward. The packaging of the sustained argument, the sustained narrative, it’s associated appurtenances that we now can see in digital form and for that matter, if it is a scholarly book, all the data sets the author wants to bring together with some opportunity to print something but not an opportunity only to imagine the print artifact as the
original version but to imagine print as one user interface that can make sense with others. I think there are exciting possibilities there. Maybe there’s somebody at this meeting already who knows something like that that’s happening. I’d love to hear it. I’m easy to find: jod@asu.edu. I think that’s the real future for somebody who will make a ton and three quarters of money selling e-books and will leave even me convinced that we have finally discovered what e-books might be. As for Western civilization, talk to me after Tuesday. Thank you.

David Durant: Thank you and good afternoon. I have been left with a lot to live up to here, so I will do what I can. My remarks are not going be so much about the actual technology, the e-book and about e-reading technology per se as about the broader impact they’ve had on our society, the possible future impact, and what we as librarians, especially as academic librarians, should take into account as we organize our own collections across all formats and future.

At the beginning of this century, e-books were, to put it bluntly, in terms of their popular appearance, there were something of a damp squib. You had wonderful projects like Project Gutenberg, and in about 2003 or 2004, I believe you had NetLibrary, which we could subscribe to and add e-books to our collection, but it wasn’t really until 2007 when the Kindle was invented that the e-book really exploded, that it took off. And so, for example, by 2011 Amazon announced that its e-book sales had actually exceeded its sales of print books. By 2012, a survey of American publishers revealed that their sales of e-books had gone from 10 million in 2008 to 457 million in 2012. So, this sort of explosion of popularity of e-books in a very, very short period of time led to what I would call the “substitution model.” You had print books. Now you have e-books. We’re going to get rid of the print books and replace them with e-books, essentially, to simplify it greatly. Essentially what happened in terms of print journals—substituting electronic journals for print journals. Substituting print reference books for electronic reference items and electronic reference databases, so this belief tended to be quite popular by about four or five years ago especially.

However, a funny thing happened on the way to our “all digital” reading future. There were voices, first of all by popular writers such as Nicholas Carr especially, as well as academics like Maryanne Wolf and Naomi Baron, a lot of concern about the impact that electronic reading and reading off e-devices as opposed to reading off of the printed page was having on our ability to engage in reading and to engage in particular in what is called linear or long-form reading, the ability to read at length, in depth, for a considerable period of time and then to be able to memorize and absorb that material and incorporate it into our pre-existing base of knowledge. There’s a lot of evidence, not just anecdotal but scholarly in many cases, that reading off of most digital devices tends to hinder, it encourages what is called tabular reading, reading short bits of information for sort of a quick piece of information here and there are as opposed to being able to read at length and in-depth, with potentially great impact on our society. In addition, the spread of e-reading and e-readers in particular, e-books has sort of plateaued in the last several years. By 2014, e-books had risen to about 30% of major publisher sales here in the US. Since then, they’ve kind of leveled off, and even I believe, according to some figures, they’ve actually dipped slightly. At the same time, print book stores last year recorded an increase in sales of print items of 2.5% here in the United States. That was the first increase in print bookstore sales since 2007. So there is a sense, a growing sense that e-books, that e-reading technology, the popularity of e-books has sort of leveled off. In particular, one phenomenon worthy of note is that e-readers, dedicated e-readers, Kindles, Nooks, their popularity has leveled off. Nearly 20 million e-readers were sold worldwide in the year 2011, according to Forrester Research. By 2014, that figure had declined to about 12 million. So, e-readers sales have declined to the point where many people are speculating that the e-readers, essentially, the dedicated e-reader, that Nook, the Kindle is headed for a sort of boutique status like the digital camera, and people are increasingly reading off of their smartphones, their tablets.

For example, the 2016 Pew Reader Survey came out just in September, has found that only 8% of people who read a book in electronic format read it off of a dedicated e-reader device. 15% read at least one e-book off of a tablet. Going back to that Pew study, the Pew studies have been done on an annual basis, and so they found some fairly consistent results that people are not generally speaking abandoning print for reading e-books. For example, the 2016 study found that 73% of people are reading books, had read at least one book in at least print or electronic format, of the total number of respondents 65%
read a book in print. 28% read at least one electronic book, and those numbers have been fairly steady. The number of people reading a least one e-book per year according to Pew has essentially been either 28 or 27% over the last three years. The number of people reading a least one book in print format, according to Pew, has essentially fluctuated from 63% to 69% to 65% this year, according to Pew’s figures. So, that figure is staying roughly the same ballpark. Generally speaking, the statistics we have show in user preferences, in media articles you see a lot of articles about millennials actually prefer to read print for long form. This is reflected in statistics, so there seems to be a general sense that, at least for now, that e-reading, e-books have sort of found their level, and that what this leads me to conclude, and leads many others to conclude, is that we should not think of a substitution model in terms of e-books versus print. We should think of a complementary model in that reading off the printed page and reading off a digital e-reading device don’t necessarily facilitate the same types of reading. They each have their uses. Each of them has their own uses, and us as academic librarians, as we build our collections and future, we need to recognize the differences between reading off the printed page and the preferences of our users in terms of reading off the printed page versus reading off a digital screen. Understand that each of these facilitates a different form of reading, generally speaking, and build our collections so that we offer our readers the best of all possible worlds in terms of reading. The ability to engage in in-depth, lengthy, linear reading off the printed page as they need to, and also the ability to engage in tabular reading, searching across text, finding brief bits of information or short passages on a particular topic off of the digital reading devices. And so, we need to facilitate both these forms of reading as we go forward. That doesn’t mean that our print collections need to stay the same size they are now. It doesn’t mean we can’t weed print materials or that we have to continue buying as many print materials as we do now, or that we can’t send print materials, at least some of them, to our remote storage facilities, but we need to make sure that we sort of retain an echo system of reading in which print and digital are coordinated, are sort of integrated to best meet the broad spectrum of the needs of our users.

In short, we need to facilitate, support what scholar Maryanne Wolf has called the biliterate brain: An ability that can both read linear, in-depth text, as well as engaging in the short tabular reading, but while it can be done in print and has been, is especially facilitated by the digital reading environment. Thank you.

Mitchell Davis: So, thank you guys all for coming out. I know the porches and decks of Charleston happy hours are calling everyone, so I appreciate you guys being here so late. I think I’ve got a little, and I think this is a great mix of people to talk about this because I’m the only one up here who isn’t a librarian and isn’t from an academic library. We’re a software company, and we license software and license content to libraries, and we work across a number of different markets, so we work in K–12 libraries. Most of our attention over the last two years has been in public libraries, which face a similar challenge as academic libraries, I think, but in a much less insulated world, and so public libraries are really having to compete toe to toe with the Amazons of the world and the media companies of the world that are just making enormous amounts of content available for very small amounts of money and really challenging what the role of the library is in a digital future. So, I bring a couple of different perspectives to that.

I’ve been coming to the Charleston Conference since 2001. I was around in that early round of reciprocal and NetLibrary and all those e-book aggregators, some of which were consolidated, some of which disappeared overnight, and did all sorts of different things. I think we really nailed it with Amazon. That’s what I was going to say is that all attempts at e-books really just had been playing at it until Amazon launched the Kindle. And even though that first device was ugly and clunky, they got it out, and the one thing they got right was it knew how to take your money and give you a book, which Amazon tends to always get right. So, and they’ve of course made the device enormously better as it has moved forward.

I talked yesterday, and I think that the past of the e-books has sort of gotten us to here, and we talked about a lot of the problems with the price of the e-books and the functionality of e-books, and for me as a startup software company in the library industry, and if any of you guys did not know, this is a harrowing industry to start a software company. Not sure if you knew that or not, but it is. It really speaks to the business structures. Most of the companies that are selling these books and these technologies are private equity-run companies.
Innovation is anathema to the entire mission of private equity. Don’t know if you knew that, but that is also true. Their job is to create consistent returns to investors, and innovation and disruption actually is a threat to that. So as long as those companies, the checks are being written to those companies, and small innovative companies are forced out of business or forced to consolidate, it is very hard to imagine how the future improves, honestly. The thing that I think is looming that I hope can sort of coalesce effort and coalesce attention is that Amazon is definitely coming for this industry, and maybe that excites universities. It definitely doesn’t excite vendors. I know that. But, it may excite universities. Who knows? I’m not making a judgment on it, but with their open education, with their OER effort, Amazon and SPIRE, make no mistake that they can suck all the profit out of this business. They could care less. They don’t care about library companies pushing each other around for market share. They will just come in and reinvent the whole thing. I think the universities spending the money really have to decide is that the future we want, or how are we going to fix this, and certainly making it easier for innovative companies to succeed I think is part and parcel of that.

One of the things that we’ve been thinking about is as we’ve sort of—we have a big project that we’ve been doing in the UK with JISC for the last two years. It’s an open education textbook project. And in that project, I think the thing that I found is that here in the US of course of the last 15 years, 20 years, the price of education has skyrocketed, the management class of universities has expanded, and it really has become more of a business. Kids are still learning things. Nobody is going to stop that from happening, but the machine of universities is a business. And I think what I see coming and what I think is good news for everyone is that in an effort to enhance student user experience, to make sure that the people writing those checks feel good about writing those checks, that the parents are happy, that the students are happy, that textbooks, overpriced textbooks, are the first thing that the university management is going to put in their crosshairs because they can improve the user experience without taking one penny out of their own pockets. So, I think that OER is going to be able to usher that in. I think that once there is enough critical mass of success with OER textbooks, I think university administration is going to get it immediately, that we can save students $2,000 a year and not lose any money. We are not selling any of these books out of our university bookstore now anyways; they’re all being ordered from Amazon and Chegg, so we don’t lose anything there. And it’s going to be a pretty radical change, and you see it coming. Because in the UK, education was free until a few years ago, and it just immediately went to 9,000 pounds a year. Most of the projects in the UK, most of the studies being done in the UK are focused on student experience because they have to justify going from free to 9,000 pounds a year. But, here in the US, it’s been more like the frog being boiled in water, you know. It’s just kind of slowly happened, and so you don’t see as much focus on that.

And so, we just got done with our first semester pilot at Liverpool. The feedback from the students was phenomenal because that’s really all we focus on. We focus on the student user experience and delivering all media types through one single interface so that students aren’t jumping out to watch a video there, an image here, a database here, and an e-book there. It all happens in the same interface. Everything is unlimited simultaneous use. There are no checkouts, returns, holds, turn-aways, so it is the kind of experience people have in their real life being brought into the university, which I think is what students expect. We’ve been very happy with that, so we are moving pretty aggressively into that.

I think the other thing that I would say just to temper a couple of things that were said before me, is that a lot of the data, and just think about this critically as you see more and more e-book data come out, and this probably relates more to the trade market than the academic market, but Amazon doesn’t give data to anyone, so any data analysis you see on e-book usage coming from traditional publishers doesn’t include 95% of the pertinent data. So, if a traditional publisher says e-book sales are flat or going down, what that means is more people are reading on Amazon, and no one can see that data. They are reading on Kindle Online Lending Library. They’re buying books directly from Amazon Publishing Imprints. One of the most fascinating sites, if you want to keep up with that e-book market, is a site called authorearnings.com where this guy has written a bot that crawls the public Amazon website and looks for changes in sales ranks and all different sorts of things. He believes he has about 85% coverage on the transactions that happen on Amazon, which is a
phenomenal accomplishment, and he every three months publishes the data on author earnings, which I also think is a massive shift, right? He’s not measuring the health of the publishing industry on publisher profits or stock price. He’s managing the health of the industry on author earnings, which is completely agnostic as to how you published your book, whether you self-published it, traditional publisher, small publisher; it’s really phenomenal. And what he is finding, it swings wildly from one three months to another, but the last report he put out was that big five author earnings in the last two years, and this is again trade, but have been cut in half, and self-published earnings have doubled in the last two years. So, there is clearly this indie revolution happening in trade, and I see a lot of parallels in that with OER that if you’ve got the right curation systems in place, if you’ve got the right peer-review, if you’ve got the right abilities to sort of push books where they need to go, that this indie revolution in a curated way can really replace these $200 textbooks pretty quickly. So, we’re very excited about OER. Our office is also right across the street, 100 Calhoun, and we’ve got an open house the next two days, so if any of you guys want to stop by and see what we’re doing with OER, we would love to have you over there. Thank you.
Hyde Park Debate Resolved: APC-Funded Open Access Is Antithetical to the Values of Librarianship

Rick Anderson, Associate Dean for Collections and Scholarly Communication, University of Utah

Alison Scott, Associate University Librarian for Collections & Scholarly Communication, University of California, Riverside

Michael Levine-Clark, Dean and Director, University of Denver Libraries

The following is a transcription of a live presentation at the 2016 Charleston Conference.

Hello, and welcome everybody to what I think is the sixth Hyde Park Debate of the Charleston Conference. My name is Rick Anderson. I’m from the University of Utah, and I will be moderating. Let me first review the structure of the debate, and then I will introduce our debaters.

Before the debate begins, the audience is going to be polled. The proposition that is under debate is that APC-funded open access is antithetical to the values of librarianship. We can do this right now. The poll is open. Each audience member is asked to cast a vote by text either in favor of the proposition or against it. Again, this is by text only. So, while I’m reviewing the structure of the debate, we’ll allow you to begin voting.

Each debater is going to open with a 10-minute statement, which has been prewritten. One will argue in favor of the proposition; the other will argue against it. The 10-minute time limit is going to be strictly enforced. After both opening statements have been made, each debater will offer a 3-minute response, and again, the time limit will be strictly enforced. Following the response, there will be a period in which the debaters will respond to questions and comments from the audience.

Following the audience comments, the audience will be polled again, and a new vote will be taken in response to the proposition. It’s important to understand that the winner of the debate will not necessarily be the one who ends up with the most votes in agreement with him or her; the winner of the debate is the one who moves the largest number of votes over to his or her side.

So, there are our debaters. First, we’ll hear from Alison Scott, who is Associate University Librarian for Collections & Scholarly Communication at the University of California, Riverside. Alison has strategic responsibility for the ways and means by which the University of California, Riverside Library’s collections grow and change. Alison joined the UCR Library in 2014, following services as Head of Collection Development for the George Washington University Libraries, Charles Warren Bibliographer for American History at Harvard University, and Head of the Popular Culture Library at Bowling Green State University. She holds a BA in English literature from Whitman College, an MLS and MA in religion from the University of Chicago, and a PhD in American studies from Boston University.

Our second debater is Michael Levine-Clark, who is Dean and Director of the University of Denver Libraries. Michael is the recipient of the 2015 HARRASOWITZ Leadership in Library Acquisitions Award. He writes and speaks regularly on strategies for improving academic library collection development practices, including the use of e-books in academic libraries, the development of demand-driven acquisition models, and implications of discovery tool implementation.

So, we will now close the initial voting for the proposition, and I’ll invite Alison Scott to the podium to make her opening statement.

Written statement from Alison Scott:

I am delighted to be here today. Whatever else I may have accomplished over the course of my career, it does appear that I have had some success as a speaker, and that I have definitely earned a reputation for having opinions.

I would appreciate it, however, if you would all take note:

I am expressing my opinion about the resolution that we are discussing as forcefully and articulately as I possibly can, but please remember that it is my opinion. My remarks here today do not represent the policy of the University of California, at the campus or the system-wide level.
So, let me begin by clarifying, to my satisfaction at least, a few of the terms that I will be using during this morning’s conversation. "Where we begin our search determines, in no small measure, what we discover."¹

By “open access,” I mean online access to published research, the materials that contribute to and constitute the scholarly record, free of charge to readers, without financial, legal, or technical barriers to access, beyond those that are “inseparable from gaining access to the internet itself.”²

By “APC-funded,” I mean the “article processing charges” that make it possible for a work—article or book—to be made available as an open access work. APCs are a means for publishers to generate the income needed to support the costs of open-access publication, enabling free access to works by imposing pre-publication fees, rather than post-publication fees, such as subscriptions or paywalls. APCs are the “author pays” tint of gold OA.

The laudable aims of open access include broadening the audience for research, maximizing the impact of research, promoting the growth of new knowledge, fostering open scholarly communication, and providing access to publicly funded research.

For scholars, [an] open access model offers the promise of increasing both the transparency and impact of their research. For the average citizen, it means unrestricted access to the published results of research financed by public funds.³

---


For librarians, open access offers practical opportunities: “Open access promises to remove both the *price barriers* and the *permission barriers* that undermine library efforts to provide access to the scholarly record,” and entices us with hope for an end to the long-standing “serials crisis.”

When I say that “APC-funded open access is antithetical to the values of librarianship,” I certainly do not mean to dispute or belittle the aims of open access publication, as ideals or as aspirations. The utopian goals of open access, in intention and, if attained, in ultimate effect, align with the many of the “core academic values and principles associated with teaching, learning and research in higher education” and the means by which academic librarians support the missions of colleges and universities. Open access is a good thing.

This is because, in part, open access, as an ideal, aligns with two of the central values of academic libraries and academic librarianship:

I believe that academic libraries are a shared resource and a community good

and

I believe that it is one of the absolutely foundational responsibilities of academic libraries, as shared resources and as a community good, to support the *creation of knowledge*—through services, infrastructure for discovery and access, collections, and all our other enterprises, by which we support teaching, learning and research.

Nonetheless, I do mean to say that APC-funded open access as a mode of publication, as it has been proposed or implemented as practice, represents an existential threat to those two ideals of academic librarianship.

---


In one of his entries in the blog, *The Scholarly Kitchen*, David Crotty wrote that

One of the core principles of Gold OA is that the costs shift from being spread broadly among consumers of the literature to being concentrated directly on producers of the literature.

This statement can be read as a bland, objective, even anodyne summary of the business principle behind author-funded models of open access.

A more critical reading, or perhaps just a more suspicious reading, suggests that this is a clear and precise statement of exactly why APC-funded open access is antithetical to the two values of academic librarianship that I have just emphasized.

First: APC-funded open access is concentrated directly on the producers of the literature.

APC or author pays mechanisms for attaining the greater good of open access mean that, practically speaking, our attention, as librarians and libraries, must be turned from support of our larger academic communities’ needs as learners, teachers, and researchers to the functional support of a much smaller group of article producers.

I will not go into the voluminous and contentious discussions about the financial sustainability of APC-funded open access—whether there’s enough money sloshing around the system, whether authors are rational economic actors, whether the subscription system can be “flipped,” etc., etc., etc.

In any case, I do *not* believe that the solution to the problem of financing open access is, as Jeffrey Mackie-Mason has phrased it, “merely” one of getting money from subscription budgets into APC budgets.” It’s not that I don’t care about about

---


money; it’s not that money doesn’t have a terrifyingly powerful impact on our work and how we enact our intentions for our work; it’s that I don’t think that money is the really important measure of the values that we, as academic librarians, need to care about.

I believe that article processing charges, under the important, laudable, altruistic guise of promoting the greater global good of the free flow of scholarly information, have the paradoxical, counter-intuitive, ironic (choose your favorite qualifier) effect of privatizing community resources.8

Second: APC-funded open access is concentrated directly on the producers of the literature.

That is, APC-funded open access revolves around the promulgation of the work products of research, and the means by which the results of research, finished works, enter the cycle of scholarly communication.

Yes, academic libraries and research libraries have an ancient and honorable responsibility to the community of scholars for the documentation of the record of scholarship, and individual academic libraries do have a responsibility to document the work produced by their own institution’s scholars.

But, in simplistic terms, I believe that academic libraries collect and make discoverable and accessible the records of scholarship—we wrangle the past—with the primary intention of promoting the use of that scholarly record for the creation of knowledge, for the sake of the future.

http://madlibbing.berkeley.edu/economic-thoughts-about-gold-open-access/

The OA 2020 Roadmap argues that

Open access cannot become a reality on a larger scale without utilizing and re-purposing the massive resources that are spent on journal subscriptions, year after year.9

In the 2015 Max Planck Digital Library Open Access Policy White Paper, “Disrupting the subscription journals’ business model for the necessary large-scale transformation to open access,” the authors stated that

[T]he final breakthrough to a comprehensive open access publishing system cannot be achieved unless library acquisition budgets are re-purposed so as to consolidate the system’s two current streams into a single undertaking to provide the best possible publishing services for the patron researchers.10

If it is true that the only way that APC-funded open access can become a comprehensive system for scholarly publication is to “re-purpose” libraries and library budgets, turning libraries away from being intentional supporters of knowledge creation and of the future, into being agents acting for a small community of knowledge producers and documenting past accomplishment, I think that libraries are in clear and present jeopardy of losing one of our fundamental reasons of existing at all.

I do not hold any brief for the wonderfulness of the subscription model of funding publication as a facet of scholarly communication, but it does have the conceptual advantage of grounding our intentionalities—and our financial conversations—on questions of the value and utility of purchased content for ongoing research, teaching, and learning.

I know I am teetering on the cusp of a slippery slope argument, but focusing on work product, and directing our efforts to managing the products of research rather than continued discovery and the future of research, even if we do have as our ultimate aim the support of the altruistic goal of open access,

http://oa2020.org/roadmap/


10 Schimmer, R., Geschuhn, K. K., & Vogler, A. (2015). Disrupting the subscription journals’ business model for the necessary large-scale transformation to open access. https://doi.org/10.17617/1.3
represents the abandonment of what I regard as two of the fundamental values of academic librarianship.

Now, as I turn you over to my opponent, the honorable gentleman from the university of the mile-high city, remember the words of Stephen Fry: “Merely because I’m expressing myself well doesn’t mean what I’m saying is untrue.”

Written statement from Michael Levine-Clark:

I am here today to argue in favor of APC-funded open access and against the resolution that “APC-funded open access is antithetical to the values of librarianship.” As I make this case, I think it is important to define some of those key values for our profession. In this context, I will start with three of Ranganathan’s five laws of library science:

1. Books are for use.
2. Every reader her/his book.
3. Save the time of the reader.

So, let’s talk about what those mean.

1. Books are for use. In Ranganathan’s conception, we can’t have books hidden away in closed stacks or chained to the shelves. If we expand that definition to modern forms of scholarly communication—to include especially online journals—publications can’t be hidden behind a firewall or accessible only at institutions that can afford the high costs of subscription. There should never be barriers to information access.

2. Every reader her/his book. We as librarians should always be able to provide our users with the materials they need, whether those are books on our shelves or scholarly articles online. We should think about information access as broadly as possible.

3. Save the time of the reader. It should be easy to access the information you need. In fact, in our online environment, it should be far easier than it ever was in the past, but to the extent that we control access via proxies and manage discovery through library-centric tools, we actually make it harder and slower to access information.

Our current system erects barriers to access and stands in the way of those three core values.

There are barriers based on affiliation. Scholars at poorer institutions (or who are unaffiliated with any institution at all) can’t access large swaths of the published scholarly literature. Even those at wealthier institutions that can afford big deals are blocked from some content, and even when their institution can get articles via interlibrary loan or document delivery, the research process slows down while they wait for access. Every reader cannot access her book (or more likely her article), and we are not saving the time of the reader.

We are forced by our license agreements with publishers to put barriers in place so that even our licensed users will have to take extra steps to access licensed content. We provide access through systems that often require multiple steps (from discovery system through a link resolver to a publisher’s website) before the user can gain access, and we force our users to log in to a proxy server or authenticate in some other way in order to validate their right to access. Articles are for use, but we make that use difficult. And again, we are definitely not saving the time of the reader.

We do need to acknowledge that article processing charges (APCs) are not perfect. In the long run, they may be just as unsustainable as the traditional subscription model. It is clear, for instance, that some research-intensive institutions would pay more for APCs than they do now for subscriptions, and a transition to APC-based open access might mean that for a period libraries will need to pay APC fees on top of their subscription expenses. APC costs are also less predictable than subscriptions and the funding sources will vary, so budgeting will be difficult.

Importantly, even though APC-funded open access will remove barriers to accessing information, moving all costs to the point of publication may well put up new barriers for some to publish. Scholars at poorer institutions, those with no institutional affiliation, or those in disciplines without significant grant funding may struggle to pay the fees required to publish. Perhaps APCs could be subsidized in some parts of the world or for some types of authors, just as there are differential subscription costs now.

But, even with those negative aspects of APC-funded open access, the net positive result is greater access to information, and that is a core library value.

As we all know, APC-funded gold open access is just one model. Another model is green open access, in which the article is published in a traditional journal, and then a version is made available, often after an embargo and often as an author manuscript, in an institutional or subject repository. Green open access does not change the current model fundamentally, nor does it remove all of the barriers to access. During the embargo period, those at poorer institutions and those with no institutional affiliation do not have access to that article at all (or at least not to the version of record). Therefore, one could argue that some forms of green open access are antithetical to the values of librarianship.

APC-funded open access, on the other hand, serves the values of librarianship

- By removing barriers to access
- By allowing all libraries, regardless of institutional wealth, to serve users, and
- By allowing users who don’t have ready access to a library to meet their information needs.

Let’s explore each of these in more depth.

1. APC-funded open access removes barriers to access. As librarians, we believe that everyone should have access to information. APC-funded open access, because it makes the article freely available to the world at the point of publication, removes all barriers to access. There are no firewalls for open access. Because open access publications are easily discoverable on the open web, users don’t have to rely on the discovery tools and access points provided by libraries.

2. APC-funded open access allows all libraries to serve users. Many of us work at institutions that can afford to subscribe to huge packages of journals, so we have direct access to large portions of the published scholarly record, and even when we can’t afford a subscription, we can generally get a copy of an article through interlibrary loan or our institution can pay for access to a PDF. We come close to fully serving our users because we can afford to.

But there are lots of libraries that do not have access to big deals. Many libraries have small enough subscription budgets that they can’t provide their users with most of the resources they need. Because scholars and students at these institutions are not able to get access to publications, their research and teaching and learning suffer. Without access to large portions of the scholarly record, faculty at these institutions are at a disadvantage in terms of being able to do cutting-edge research, secure grants, and get published, and because they can’t get access to the latest research, their teaching may suffer too. Students at these schools are at a disadvantage in terms of learning outcomes.

But with access to articles funded by APCs, all faculty and all students can get the resources they need to grow as scholars and teachers and learners, allowing them to be competitive with students and faculty at information-rich institutions.

3. APC-funded open access allows users who don’t have ready access to a library to meet their information needs. Just about everyone in this room has access to an academic library that subscribes to at least the basic resources they need, and even when those resources aren’t enough, our libraries will secure us additional resources through interlibrary loan, document delivery, or even a new subscription. Even the poorest libraries can do at least some of that.

But there are unaffiliated researchers all over the world. Some of them are even our alumni. Who here hasn’t had to tell an alum that she would no longer have access to the resources she became used to while studying at their institution? Our current system means that people who want to conduct research professionally or learn about something new for personal needs are cut off from most publications if they are not affiliated with an institution that
can cover subscription costs, whether that institution is a university or a think tank or a corporate research environment.

Open access solves this problem. Green open access breaks down those barriers by giving post-embargo or author manuscript access to people who are not affiliated with an institution. APC-based gold open access gives them immediate and direct access to the version of record. With a complete transition to APC-based gold open access, we all would have equal access to published scholarly research.

In summary, while there are clearly some flaws to APC-based open access, most notably that APCs erect a barrier to publication, there are clear benefits. APC-based open access provides greater access to information, something that we all should believe in. APC-based open access supports three key library values:

1. That publications are for use, that there should be no barriers to access;
2. That every reader should have access to his/her publications, that all publications should be accessible to all readers; and
3. That we should save the time of the reader, with no extra steps between discovery and access.

Article processing charges allow us to serve our users better and are definitely not antithetical to the values of librarianship.

Response From Alison Scott

I’ve always wanted to bang my shoe on a lectern and shout, “Of all the damn nonsense!” But I actually agree with Michael for some of his key points. Open access offers us great promise that barriers to information access will fall, that information access will no longer depend on location or affiliation, and that, when the Jubilee comes, access to information will be easier and faster. Further, my learned opponent warms my heart with his foundational appeals to three of S. R. Ranganathan’s five laws, although I think he also could have included the fourth law, “Every book it’s reader.” This law seems just as pertinent to his argument as the other three so far as all four of them keep our attention centered on readers and not on producers. However, I have to take issue with one of Michael’s examples of historical barriers to access. Chained books should not be simply dismissed as the barbarous invention of a barbarous age. In times of desperate bibliographic scarcity, chaining books to library shelves meant that thieves, or just the selfish, could not privatize the common good of texts that were meant to be shared by a community of readers. By pushing back against this specific example, I don’t mean to equate APC-funded open access with book theft, and heaven knows we do not live in an age of information scarcity. But I do repeat my charge that article processing charges privatize the resources and the intentions meant to support the library as a shared community resource, even if the ultimate altruistic intention is barrier-free access to information for a global community. Essentially, Michael’s argument is that the end of open access justifies the means by which we attain it. I am reminded, as no doubt you all are, of Mary Wollstonecraft’s comments on the origin and progress of the French Revolution. “Malevolence has been gratified by the errors they have committed, attributing that imperfection to the theory they adopted, which was applicable only to the folly of their practice.” Open access obeys four of Ranganathan’s laws, but I hope I am not being malevolent when I say that the author pays costs model of funding is, in Wollstonecraft’s terms, error, imperfection, and folly.

Response From Michael Levine-Clark

So, my remarks aren’t going to be as funny. Sorry. Open access as an ideal is fundamentally a good thing, and clearly, we both agree on that. It does align with the values of librarianship as we’ve both described them. But I’d like to dig into the values that Alison articulates, values that I also agree with, by the way, and talk a bit more about why I do not see APC-based open access as a challenge to them. She states that, “Academic libraries are a shared resource and a community good.” But our current model of subscription funding and license access is fundamentally at odds with that value. A typical research library, again, spends millions of dollars annually on subscriptions that can’t be shared beyond the licensed campus user base. The current subscription model makes us less and less a shared resource every year as larger and larger portions of our collections end up behind firewalls. She also states that academic libraries, “... support the
creation of knowledge.” And honestly, I believe that if we’re not doing that, then our universities probably need to shut our doors, but I don’t see how we can support knowledge creation without providing access to the published literature that supports future research, discovery, and learning. With subscription-based access to that material, we reinforce a system of information have and information have-nots, and those have-nots are at a disadvantage of creating knowledge, but APC-funded open access decreases that disadvantage. More students, more faculty, more people generally have access to these publications that serve as building blocks to support knowledge creation. APC-funded open access means that libraries can invest in discovery, services, and spaces rather than collections, and it means that all academic libraries can support knowledge creation.

Alison presents us with a choice between, “. . . focusing on work product and directing our efforts to managing the products of research,” and on helping with, “. . . continued discovery and the future of research.” I believe we as institutions of higher learning can and should be doing both of those things. To some extent, though, I believe that the day-to-day management of compliance in APC funding is not so much a library function as a function of the office of research. If that’s true, then it frees up the library to focus more on curation, access, discovery, and service. While I am deeply concerned that a switch to APC-funded open access could impose barriers to publication for some authors and that will hit certain disciplines and certain institutions harder than others, I also believe that APCs will generally allow greater access to information for all. APC-based open access should make published research more accessible to more people through more libraries, allowing us to focus on that fundamental task of facilitating knowledge creation.

![Figure 2. Closing poll results.](image)
Following audience questions and comments, the audience was asked to participate in a closing poll.

Rick Anderson: Leah, if we could bring up the poll again. And again, I would invite everybody; this works by text only. Please register your vote either in support of the proposition or against the proposition. In the past, we’ve said we’ll give you 5 minutes, and then after 2 minutes the voting has kind stopped, so, . . . And just for your reference, the opening poll results were 54 in favor of the proposition and 124 against the proposition. This is a little confusing, when I say in favor of the proposition that APCs are bad. So, against APC’s, therefore in favor of the proposition. Looks like we still have some votes coming in. While the votes continue, Michael will sing.

Michael Levine-Clark: You really don’t want that to happen.

Rick Anderson: It might sway the voting inappropriately, ha-ha. (Lengthy pause while votes are cast.)

Rick Anderson: It is looking like a clear victory for Alison Scott. Congratulations, Alison!
Working in Partnership to Support Quality Research

Jayne Marks, VP of Global Publishing, Wolters Kluwer

The following is a transcription of a live presentation at the 2016 Charleston Conference.

Jayne Marks: Good morning, everyone. So, before I get started, first of all, I should say thank you so much for inviting me here and having me come talk at the Charleston Conference. I have to say this is my absolute favorite meeting of the year. I really enjoy it. It is so great to come and meet with people who really get what we do, all of us, the publishers and librarians working together talking about real issues. That is what I really like.

How many in this room, and I want you to be really honest, think that publishers are just after building pay walls and making money? Come on! Three of you? Okay. All right, so I’m in for an easy ride. Now, I know that is a lot of the public perception of what publishers do today. I have been in journal publishing now, I did admit to my colleagues last night, I’ve been in journal publishing 35 years. I tried a couple of times to get out. I keep being drawn back in. I love publishing. I love academic publishing. I’m now in medical publishing, so I work exclusively in the medical space, and I’m very passionate about this space, but I know that we have a huge image issue. So, what I want to do today is to share with you some of my thoughts about how really at the grassroots, on the ground we work together, and we work together really well, and we support our community of interest.

I don’t know how many of you were at James Neal’s room, and I want you to be really honest, think that publishers are just after building pay walls and making money? Come on! Three of you? Okay. All right, so I’m in for an easy ride. Now, I know that is a lot of the public perception of what publishers do today. I have been in journal publishing now, I did admit to my colleagues last night, I’ve been in journal publishing 35 years. I tried a couple of times to get out. I keep being drawn back in. I love publishing. I love academic publishing. I’m now in medical publishing, so I work exclusively in the medical space, and I’m very passionate about this space, but I know that we have a huge image issue. So, what I want to do today is to share with you some of my thoughts about how really at the grassroots, on the ground we work together, and we work together really well, and we support our community of interest.

I don’t know how many of you were at James Neal’s session just a little while ago; he said two things that really struck me. One I completely agree with that, and one I completely disagree with, so they might surprise you. The first one that I agree with is he said, “We are in a state of constant change.” I don’t think anyone would disagree with that, whether you’re in publishing or academia or librarianship. Everywhere so much is changing so fast that really change is the new normal. We have to get used to that. But, the other thing he said was, right at the beginning, he said, “The community of interest between publishers and librarians is narrow.” I’m not sure I can really agree with that because everything that we do in publishing space is for your patrons. That is what we do it for. We do it for your patrons to make their work out there in the world, to help validate their work, to disseminate their work, to make sure that it is preserved long-term, and I hope that in those senses we have the same community of interest that you do.

Having set that as my context, I want to talk through, well, let me start with my agenda here. Talking a little bit about actually creating and delivering that quality of research to the finished product and then how do we get it out to the widest possible audience and what do we bring to the table? What do you bring to the table? How do we work together? And what I would really like at the end of this session, after I’ve given my thoughts, is to hear your thoughts on what we could do more of and how we can help each other better.

So, let’s start with delivering quality research, and let’s start with the authors. The authors are the most important people that we work with in terms of the scholarly research continuum. The authors and the researchers are the people that we need to support the most, and I think that the other thing that James Neal said that I completely agree with is the complexity that researchers are facing today in terms of accessing and publishing their work is really, really complex and hard. And I think the thing where we work best together, and we do a lot of together, is training information, making sure particularly that young authors as they come into the scholarly communication process understand the complexities, understand the choices, there are lots of choices out there. We deliver webinars. We deliver a lot of online information resources, in-person training at a lot of conferences. A lot of what my publishers do who manage journals in our organization, when we go to meetings of our society partners, they’ll run sessions on how to get published, how to make sure your work is in the best possible state, where to go and get help if you need help, and how to publish your work. Then we provide support services. I’m sure you do the same thing too. We provide language editing. There are some great resources out there, one of which we partner with, I know other publishers partner with the same, that you can point authors to who maybe haven’t published before, or English is not their first
language, and maybe even the scholarly process is different in their country, and they need help in understanding what makes a good piece of research and a good publishable paper.

We provide training courses on the ground in multiple countries to help people to do that. In fact, we have an organization that is part of ours in India that supports emerging markets publishing. If you imagine you are a Nigerian physician, you’re doing work on the ground in Nigeria. You want to publish a clinical study that you’ve done. It’s not going to be of interest to a Western journal where the American Journal of, I don’t know, say Ophthalmology, has all of the tools and equipment you could possibly imagine to treat a patient, that Nigerian ophthalmologist won’t have those. So, you need to be able to understand and to put that in context of the work that he is working in. But the peer review has to be good, and the process has to be the same; it just has to be in context. So, we provide training courses for authors in those countries, and we provide training courses for new editors who have perhaps just been tapped on the shoulder and said perhaps you’re going to be the new editor of the new journal of Nigerian Journal of Ophthalmology, and they have no idea what to do. So, we provide a lot of those resources, training, there’s some great collaborative resources across publishers and librarians to help do that.

And then mentoring young researchers, in each of the publishing associations there are mentoring programs for young publishers. I know there are mentoring programs in a number of different areas for young researchers, a number of our partner societies who own the journals that we publish have mentoring programs for young peer reviewers. For example, there might be a special program that they are able to sign up for, and they can learn how to do peer-review. It is not something you just grow up learning how to do. It’s all about training, and it’s all about helping us to get to better authors because better authors definitely gives us higher quality of research.

I wanted to put this up as an example. There are lots and lots of examples out there, but I found this one, and I thought this was great. Here there are links to all kinds of resources that the library, this is a library site, that this librarian has put together to show authors where to go and more importantly down here where not to go. What are journals that perhaps they should be a little bit more careful of? Because I think that’s one of the things that we can work together to do is to really help authors to understand when perhaps that offer of a publication isn’t quite what they expect it to be. So, that’s the other thing we can do. We can help authors navigate this maze of where to publish. We’ve helped them get their paper into a publishable format. Where are they going to publish it? What is most important to them? Is it prestige? Is it speed? Do they want open access? Is open access important to them? And how is that going to help them? All of those things I think together we can help researchers, particularly young researchers who are publishing their early papers, to understand what the options are, where they should go for-publishing their first paper in Nature or Science is probably shooting for the moon. It’s probably not going to happen, so where do you start? Where do you want to put your paper? And when is it really important to get the speed and where do you go? There are journals that specialize in almost every area of this, particularly including open access, and therefore helping authors to know where to go to navigate that landscape and to understand the benefits, the upsides and downsides of that, I think is really important. If you are a tenured professor publishing your 200th paper, you’re unlikely to be worried, perhaps, about whether or not the journal has an impact factor. Maybe you just want speed, and maybe you just want to get that out in the market, and you know that you can get into a journal that’s going to get it out fast. But if you’re an author on the tenure track, getting into the right journals, because that’s the system that we work in, whether we like it or not it, that is important to them.

The last question I think is really important: Is the journal authentic? I want to show you an example that came up, and this an example that came up very recently. So, this author got this e-mail, “Would you like to submit in Plastic and Reconstructive Surgery?” The e-mail goes on. This is sort of the follow-up. You notice that by the time it gets into the e-mail it says Gavin Journal of Plastic and Reconstructive Surgery, but the headline is “Plastic and Reconstructive Surgery.” Then you go to the website, and the website is suddenly “Plastic Surgery and Modern Techniques.” So, what happened there? The real journal is this one, and this is one we publish, and the society came to us and said, “What do we do? There’s this plastic and reconstructive surgery journal out there that’s not one of ours and is
masquerading as our journal.” I cannot tell you how many times we have people come to us, day in and day out, saying I found the journal of this, or the journal of that, that is pretending to be our journal. They're very clever. They have covers that look like the covers of the journals. They present themselves in very similar ways, and they do things like this with the e-mail. It is very easy to be taken in by some of these if you are new into the field. And I think that this is an area where we need to do more to help people understand when an offer to publish is really just a Nigerian princess e-mail, and they should avoid it. There’s a wealth of difference between pure predatory publishing and sort of the broad range of open access journals. There’s lots and lots of different journals in between, but there are clearly predatory practices out there that we can help protect our authors from, and I think we can work together on that. How many of you have heard of the Coalition for Responsible Publication Resources? Good. Somebody’s heard of them. So, this, I just highlight here is it is a conglomeration of actually Don Samulack from the Editage Company is working with a group of publishers at the moment to try and put together some resources that will really help people understand when predatory publishing is happening whether it’s bad practices in peer review, whether it’s bad practices potentially in submission, but also just journals masquerading as journals that really are not bonafide journals. So, I think this is an area where we definitely can work together more on.

Peer review is really important, and I’m going to go through these roles. I’m sure you know these roles as well as I do, but they are complex. They’ve actually become in some ways more complex. Authors clearly, they write their paper, they write up their results, and they submit their paper. I’ve put under here interestingly who archives their paper? Because many of the funders put the responsibility for the archiving on the author, but I suspect that many of us in this room would say actually it is librarian that archives it, or publishers would say, “Actually we’re responsible for depositing that paper when it is ready.” So, I think really archiving the paper is somewhat of a shared responsibility between all of us. Manages data is an interesting one, and I want to go back to that one later. The editor sets the editorial policy, whatever that journal’s editorial policy might be, appoints an editorial board to help him or her to do their job, and it might not just be one editor. It may be multiple editors. They choose the reviewers, and generally speaking, they will make the final decision, and at the very least, they will make the final decision on any potentially controversial papers. The editorial board, on the other hand, they provide a lot of support. They usually do a lot of the reviewing, and they do a lot of promoting of the Journal and helping to get authors to come in. Sometimes different editorial boards will be split into different subject areas, and certain subeditor’s or groups will take different responsibilities. Reviewers, they are so crucial, and this is an area where we don’t, in many cases, we don’t do enough to support editors and reviewers in their work. It is really a crucial role of the reviewer to assess the accuracy, to make sure that the data that is in the paper supports the conclusions and puts that context or the content of what is being published in the context of the broader world and then matches that to the journal’s policies, whatever that journal might be. So, it might be a highly selective journal. It might be a broad-based journal, whatever it is, or they might have published an article on the same topic last month in which case they might not be so interested. So, that’s the reviewer’s role.

What does the publisher do? They provide systems and infrastructure to manage this process in terms of the technology. They, in most cases, fund the editorial office, and the editor might well have an entire team of people that backs him or her up in terms of managing this process depending on the size of the journal and providing support in terms of training, training on how to use the system, bug fixing when the system goes wrong, support for the authors in terms of doing that. More and more I think publishers are finding ways to help thank reviewers for what they’re doing, maybe sometimes offering access to content. In our world, we offer reviewers CME credits, which is, particularly for somebody trying to amass their credits for the end of the year, it is important to them to get those CME credits. So, finding ways to help this process, support this process is something we can definitely work together on.

Archiving is incredibly important. I know it is the key role of the librarian to make sure that archiving happens, but I think it is something that we share in in making sure that it works. There are so many different archiving policies out there. Funders have different policies on how things are published, where they are deposited; it is amazingly complex.
I’m sure from your perspective, but certainly from a publisher perspective, with papers coming in from all over the world under many, many different funder mandates, it is bewildering to keep up with it. We provide, between us, you provide the institutional repositories; we provide the help in terms of depositing in archives, depositing directly into a number of funder repositories, including PubMed Central is probably the most important one. So, in that sense, actually making sure that that process happens and that we support the authors together and making sure that that process happens is important.

I’ve put up here creating and managing a data management plan. We spent a lot of time scratching our heads within Wolters Kluwer with what does data mean in terms of how do we support it within the publishing process? And this is a personal view of mine; it is not necessarily a WK view. I think we are still thinking about it, but I don’t think it is the publishers’ role to preserve the data. I think there are services out there that do that. I don’t know if it is the librarian’s role to manage the data and the data curation and the data sort of storage going forward, I’d be really interested to hear your view. Is it the role of, for example, the NIH? Is that the place where data should be deposited and curated? I think having a data management plan is a really good idea. How that happens is something that I’m very interested in and still not convinced that we have all the answers at all.

And then the last thing, providing archiving services. I think every publisher deposits their content into one or more of the archiving repositories to make sure that we have that long-term preservation and most importantly the long-term preservation of the version of record so that as the record evolves, as authors come back and add or change or amend their paper, maybe not amend but add to it more information, then we need to make sure that all of that is captured and captured for the long-term.

So, that is how we get papers. Now how do we get it out into the world? How do we get in the hands of people who need it? Finding content. This poor man has got his head in the haystack. James Neal was absolutely right: Trying to find the right content in today’s world is incredibly difficult. Of course, everybody starts with Google Scholar, you hope, or Google. You hope they go on to the library services, use the discovery services. Often they use social sharing networks to find the paper they’re looking for or to just find papers. I think there are different ways of searching for something depending on whether the researcher knows what they’re looking for or just trying to look at the landscape of maybe a new area they want to research. I think there’s a lot of different ways to search for content, and I’m not sure we’ve all quite got it right yet. There are a lot of tools out there, and I think we need to make sure that researchers on the ground know how the different tools work and how they can use them more effectively. How do we help with this process? Metadata consistency, and did I see Todd come in the room? Todd Carpenter? If he’s not here, he certainly is really keen on metadata and standards. That is something that we as an industry, in the publishing industry, have to work really, really hard to maintain, because if we don’t have consistent standards, and you can’t pull that content into your discovery services or other systems that you want to work on. I understand there is a new JATS. First of all, we had to amend everything to take in JATS. Now there is a new JATS, which is the DTD format, so everything has to change. The standards in the industry are crucial, ORCID for understanding who the researcher is and being able to de-dupe researchers. FundRef so that you can as a funder actually understand where the research that you have funded finally gets published. The DOI, just the basic DOI, is critical and then how we use the DOI in linking through CrossRef and CHORUS. How many of you are aware of the CHORUS initiative? Hopefully most of you. CHORUS uses a lot of these standards to help different funders to actually get to the papers that they have helped to fund, so all of those things I think together will continue to evolve. I think we need to do a lot more of that, and it is something that we absolutely have to work on together to make sure that we’re really marching in the same direction. And then I mentioned earlier the version of record. I think this is really important, and I think it often gets forgotten in the myriad of ways that people can get to content. Maybe I’m more sensitive to it because in health I think it is really important that the version of record is the one that a physician might use to decide to change their practice, and they need to know whether there’ve been any updates to that version of record.

Next, discovery services. I’m sure there are more than one other session here, and I know that there is at least one other session at this meeting, and there is usually multiple on discovery services. It is pretty
much a Space Odyssey. There’s so much content out there, and there are a number of different discovery services all trying to help you and us and our patrons, our joint patrons, our readers get to content they need. We need to make sure that the solutions are workable, usable, they’re transparent, that we partner with them, and that they’re trusted. For example, it is really important that a discovery service gives you access to everything and doesn’t imply any kind of preference on pulling anybody’s content to the top. It has to be what the researcher is trying to find. Linking and data, I mean how are we going to then evolve discovery services to take in all these data repositories that are going to be out there? I can’t imagine what that is going to be like.

So, actually, and then providing metrics and actually providing measures back to the library about what that content is that’s being used. It was interesting that James Neal said, “I’m not going to pay for anything that my patrons don’t use.” I completely understand that. How do we make sure that when a library patron needs some content that you have that they don’t go out to Sci-Hub, for example, to get it because it is easier? How do we make sure that together we work to make it easy to get to content, and I think that is something that we’re not doing as well today as we could be.

Really that comes down to the last point: How do we get content together? How do we make sure that content gets into the hands of the users in their workflow? And that is why I do completely agree with James Neal when he says it’s workflow tools that are important, and I think it is something that publishers are focused on, I’m sure it is something that you as librarians are focused on as well. So, here’s a couple of examples. You’ve all seen different discovery services. We actually just launched one this year for the hospital space because many—in fact many hospitals don’t even have a library. The hospitals who don’t have a library really need some way of pulling all their content together so there are different interfaces available for different markets. So, then how do we measure? How do we measure the quality and the usage of the content that we all publish? It is as important to us to know that the content that we are publishing is being used. It is of no interest to me to publish reams and reams of articles that nobody ever reads. The important thing is to get those articles out there and to make sure they’re being used. There’s lots of ways to do that and we all know Scopus, ISI, Altmetrics; we need to make sure that our users, and dare I say that maybe even senior people within publishing and senior people within the academic sector, can understand what these metrics mean and most importantly understand what they don’t mean. Just because something has an impact factor does that mean that if the usage in your libraries is low is that good or bad? If one person gets to the article they really need to read and maybe it has a high altmetric score, is that good or bad? I think this is something that we haven’t figured out yet, and I think we are still learning about how to navigate this sort of maze of different analytics that are coming back to us, and I think that we need to understand how these things are telling us how our patrons want to use the content that we make available to them.

I want to talk a little bit about marketing because that’s something that the publishers do, and it’s very focused these days on getting the content to the right people. It’s much, much more difficult today than it used to be, and this is a slide that I borrowed from my colleague who heads up our marketing department, and he’s really talking about how marketing has evolved. Marketing is no longer putting a message down on a piece of paper and mailing it to somebody. We’re long past that, so people working in our marketing department had to have these kinds of skills or these kinds of tools to be able to do their jobs better. Analytics, first and foremost, is critical. I do not want to send anybody an e-mail that is of no interest to them because that is wasting their time, and it’s wasting our time. So, really understanding the relevance of the content that you have, the message that you have to the audience is critical, and only by using analytics can you figure that out. Social media, I don’t need to tell this audience that social media is really an important method of communication, and how we can use that effectively and reasonably to get again content being discussed out there in the market is important. Once you have all this data back, can we use data visualization more to actually be able to show us what is important, what’s happening in the content that we are publishing and that we’re accessing? Marketing people don’t have to be just creative anymore. They have to have technical skills or least within your team, as in the next one, you had to have technical people as well as people who are good and creative. It has to be a team approach today. This one is interesting. I’ve never heard of “newsjacking” but actually taking advantage of when something is in the news of saying, “Hey, I know this is in the news today,” that here in this academic
environment we have some fantastic content that really sets the context of the news that you’re reading today. So, I mean I’m sure as librarians you’ve had people come in and say, “I’ve heard about the Zika virus.” Well, probably not recently, but maybe when it first came out. “Can you tell me about it?” If we as publishers can get ahead of that and give you content sets or give our readers content sets and links to information that is really important to them, I think that is going to—that is something that we can do as a service, and I think that it is going to be very much more helpful, putting information in the hands of people who need it. And soft skills. I think everybody needs that, but here we most definitely within a marketing team today they have to be able to work well together. So, that’s in that sense, we use all of those skills to try and get the right information out to the right people.

So, to sum up, I don’t know how I’m doing for time, a little bit over, so, this is my summary, really. We can and we do work together to support authors. They are our community of interest, and I think we can and we already provide them information training, guidance. We can do more, and we’re always interested in hearing from our library partners on what more we can do there. Preserving and nurturing peer-review is critical. Lots and lots of much, much smarter people than I talk about whether or not peer review is no longer needed, can be superseded, as far as I can see today peer-reviewed is critical at being able to say, “This paper is something you can trust.” And let me tell you, in the medical field, you want to be able to look a physician in the eye and say, “This paper is something you can trust,” because if they’re going to go out tomorrow and do something different with the patient, you want to be sure that paper is right. I do think we need to make sure that we work together to support our readers and patrons to be aware of what information is out there and what information is trusted information and maybe where they should be a little bit more careful. And lastly, I would just leave with this thought—well I have one more thought, but leave with this. We are in a complex and dynamic environment, and things are changing very fast, and we have to continue to do what we are over these three days, talking to each other in detail about the systems we have, the processes we have, the challenges we face to make sure that we can make our patrons’ and our customers’ lives better. I’m going to leave you with one last thought, particularly as the editor of this book is in the room, this is a shameless plug for a new book that’s coming out toward the end of this month. It is all about academic publishing. It’s written for librarians, very conversational style. I should probably own up to the fact that I have a chapter in this book, but, if you’re interested, it is published by Mission Bell Media, and it’s coming out at the end of November. I highly recommend it, of course, because I’m biased. Thank you. So, questions?
Libraries as Convener, Enabler, Distributor, Advocate, and Archive in the Future Knowledge Economy

James G. Neal, University Librarian Emeritus, Columbia University

The following is a transcription of a live presentation at the 2016 Charleston Conference.

James Neal: It’s an honor to be here at the Charleston Conference, and I was here several years ago to give a presentation, and I applaud the extraordinary people who have made this conference possible over many, many, many years.

I have noticed over the last several years that my presentations at professional meetings have become much more alarmist and much more strident. Maybe that’s a reflection of retirement. I have subscribed to the Emerson adage that sometimes a scream is better than a thesis. Prognostic exercises offer opportunities to set aside reason, to avoid evidence, and to speculate with abandon. Library futures and perhaps publisher futures are particularly challenging to define as the community of interest is narrow, and the implications of error are modest. As Ken Kesey, author of One Flew Over the Cuckoo’s Nest, once remarked, “You can count the seeds in the apple, but you cannot count the apples in the seed.”

I went to Columbia in the fall of 2001, and I agreed to give a series of presentations to alumni groups around the mid-Atlantic region. That January I found myself driving down the New Jersey Turnpike in the snow to speak to the Eastern Pennsylvania Columbia Club. I went over the bridge, into the city, found the hotel, parked the car, went into the hotel, found the room. My talk was to begin at 8:00. By 8:15, there was only one person in the room. I suggested that I go ahead and give my talk and he said, “That’s great.” I was using slides, and I asked him if he would show them as I spoke. He did. I finished my talk, actually it was a really good one that night, and I asked if there were any questions, and he said, “No, I have no questions.” So, I sat in the front row, and I said you know, I’ve got to get out of here. I’ve got to get back to New York. And he said “No, no, you’re going to stay.” I said “No, I’ve got lots of meetings in the morning, and you’ve got snow on the road, and I’ve got to get out of here.” “No, no, no, you’re going to stay.” I said “You don’t get it. I’ve got to get back to New York tonight!” He said, “No, no, no, you don’t get it. I’m the second speaker.” So, I always love it when I’m the second speaker.

Libraries have entered a period of gross mutability, a state of constant change, of productive and powerful chaos, of hybrid strategies and maverick structures, of radical shifts in professional staffing, of massive leadership turnover, and of essential creativity in advancing our individual and our collective visions. There are, in my view, three essential elements. First, we must have hope. Believe in and aspire to expanding relevance and impact in the communities that we serve. Second, we must achieve power to have authority, influence, and respect. And third, we must focus less on ideas and more on action, getting things done. The two things we must advance are primal innovation, a basic commitment to risk and experimentation, and radical collaboration, deep and systemic partnerships. Renovation is grossly inadequate. Deconstruction is totally essential. This means redefining the physical, the “where,” the expertise, the “who,” and the intellectual, the “why?” Infrastructure of our libraries and understanding the psychology, the economics and the methodologies of progress. Progress. Samuel Butler tells us that all progress is based on a universal innate desire on the part of an organism to live beyond its means. George Santayana points out that those who speak of progress measure it by quantity and not by quality, and Khalil Gibran points the way progress lies not in enhancing what is but advancing toward what will be.

The library has always been a fundamental partner in the learning and research processes, but key changes in the information technology, economic, social, and political environments are challenging this relationship and raising critical questions about the value and impact of the library in the community. Do 20th century skills still matter? The work of information selection, acquisition and synthesis, the support provided for navigation, dissemination, interpretation, and understanding, the tools for use, application, and archiving of information—does the community still need the support in the ways that we as libraries have

Copyright of this contribution remains in the name of the author(s).
https://doi.org/10.5703/1288284316499
provided them over the last 50 years? And do the new roles that libraries are advancing as aggressive consumers, intermediaries and aggregators, publishers and educators, research and development organizations, creative and maker spaces, entrepreneurs and policy advocates, do these present a refreshed opportunity for innovation and library centrality in the university and the community? For me, it means that the library must be virtual, engaged with users in evermore rigorous and effective ways, in the classroom, in the laboratory, at the workplace, at the hospital bedside. We also must be virtuoso, smart but ready to learn, expert but always compassionate. And we must be virtuous, radically partnering and always working in the interest of our publics.

This brings me to my two main theses for my presentation today. As we look out over the next decade, libraries will be increasingly defined as convener, enabler, distributor, advocate, and archive and less as infrastructure, platform, repository, and portal. I also propose that by 2026 there will be no information and no service industry targeting products to the library marketplace. Let me read that again. There will be no information and services industry targeting products to the library marketplace. Content and applications will be directed to the consumer. Open resources for learning, research, and recreation, and open source tools supporting individual and organizational productivity and innovation will be much more prevalent in the global economy. Self-publishing and niche technology development will dominate. Information policy wars will dictate national and global legal and legislative debates. Libraries must be effectively integrated into new creative environments. Libraries will systematically apply new knowledge to new resources to produce new goods and new services. That is we will be much more focused on developing the market. We will focus on managing the costs and increasing the benefits. That is, we will find ways to add value. We will think deliberately about existing challenges and unmet needs. That is, we will focus on solutions, market value solutions. We will understand the importance of achieving balance between evolutionary, that is, incremental change, and revolutionary, disruptive change. We will go through measured transformation.

What do I mean by transform? To change in composition or structure what we are and what we do. To change in outward appearance or form how we are viewed and how we are understood. To change in character or condition, how we do it. Thomas Kuhn in The Structure of Scientific Revolutions reminds us that the transition from a paradigm in crisis to a new one from which a new tradition can emerge is far from a cumulative process. Karl Marx in his theory of epistemology, his theory of knowledge, talks about a pot of water over a flame, and intellectually we know that the temperature of that water is increasing, but only when it reaches a certain point, a boiling point, a tipping point, does a true transformation take place as the liquid turns to gas. The fundamental link between a cumulated, quantitative change and qualitative change.

I worked at Penn State University for seven years back in the ‘80s, and I remember a wonderful story about a physics professor at that university who decided to climb this mountain next to campus. We called it “Mt. Nittany,” and he wanted to explore with the gods who lived on the top of that mountain whether the work that he’d been doing his entire life to discover a universal theory of matter would actually occur. He climbed to the mountain and confronted the gods with that question. “Will there be a universal theory of matter?” And God thought briefly and came back and said, “Yes, professor, there will be a universal theory of matter but not in your lifetime.” Well, that was not too bad because all of his work was actually going to pay off into something important. Word circulated around campus about this faculty member’s experience, and the president of that university decided that she also should climb to the top of that mountain and confront God with a problem had been plaguing her colleagues around the country for years. Reaching the top she asked the gods, “Will the cost of libraries and technology ever come under control at my university?” And the gods went off and thought and thought and thought and several days later came back and said, “Ms. President, yes. The costs of libraries and technology will come under control at your University but not in my lifetime.”

Let’s consider for a moment where libraries actually spend their money. There are four main buckets: Content, purchase or rental, technology (systems, applications, tools), staff expertise and space. Clearly, investment in most libraries, in my experience, is flowing from content to expanded investment in these other areas. Libraries are fundamentally rethinking space planning and identity. We are creating the trompe l’oeil library
with many of the superficial trappings of the traditional library, but with expanded understanding of user needs, user expectations and with technology as a catalyst, we are building learning spaces, social spaces, collaborative spaces, creative spaces, all defined by flexibility and adaptability. It’s going to take more resources to do this well.

Libraries are experiencing a rapid shift in their staffing. We’re seeing more professionals with more diverse academic and professional backgrounds. We are recruiting for a wide range of new professional assignments as the role and responsibilities of the library have expanded. We’re seeing more messy and more fluid organizational structures which require a new style of leadership, and we are striving to build organizations characterized by diversity and inclusion to reflect more the communities that we serve. This means more competition for successful recruitment and retention of staff. It’s going to take more resources to do that.

Libraries encompass and espouse technological change, often serving as the locus of early adoption in our communities. We have built digital libraries, recognizing that in doing so quality equals content plus functionality. It’s not just providing the stuff. It’s enabling people to use it and apply it in their work environments. Multimedia, integrated services, and applications are central to the digital future. We have been so focused on our library management systems. Yes, the need for inventory control persists, but many of us have created new discovery and access tools to support our users. We have built those front ends, but the real action is in new technologies and platforms, mobiles and tablets, cloud computing, Geo everything, personal web, artificial intelligence, linked data, big data, semantical ware applications, smart objects and smart spaces, open learning, games as learning tools, visualization and simulation, 3-D printing, augmented reality. It is going to take a lot more resources to incorporate these and future technologies into the information service programs of our libraries, but it’s important to keep in mind as we do so: The early bird may get the worm, but it is the second mouse, the second mouse that gets the cheese.

I recall the wonderful Mel Brooks film, “History of the World, Part 1.” How many of you remember that film? Well, there’s a great scene when Brooks, as Moses, is coming down the mountain carrying three large stone tablets. “Children of Israel, I have 15 . . .” He suddenly trips, and one of the tablets crashes to the ground and falls apart. He picks himself up and proceeds down the mountain, “Children of Israel, I have 10 Commandments!” I think we all applaud the loss of those five additional rules, but allow me to provide some speculation as to what they might have been:

Thou shalt preserve the cultural and scientific record. Moses was really smart. We, libraries and publishers, we’re in deep trouble. The world is producing vast amounts of digitized and born digital content. The volume, complexity, and dynamism of this information challenge forces us to think creatively about its capture, its organization, and its long-term preservation and usability. Internet pioneer Vint Cerf warns us about the risk of a digital Dark Age. If we do not develop the technologies, the tools, the financial resources, and the shared responsibilities to address the risks to our cultural, scientific, societal, and community records, we are in trouble. We have done a very modest job at best in preserving the analog record. We are failing in our management of the digitized records, including e-journals, e-books, e-media, and e-documents. And for born digital materials, although we see pockets of activity around the world, there are minimal sustained programs and investments being made. This is an issue of integrity. We must maintain human records as complete, unimpaired, and undivided as possible. The ability to consult the evidence and sources used by a researcher and author will be lost if those digital records are not available. If I can’t look at the born digital sites and footnotes in your paper, then I have to question what you wrote and its accuracy and validity. The ability to research and investigate the history and current state of our world will be compromised if born digital materials are gone or changed. The ability to assess the sources of record will be very difficult if they are deposited and dispersed as they are into multiple and disparate sites. This is the challenge of repository chaos.

At the core of digital preservation, for me, there are four principles. We hold the content, the archive as repository because we cannot preserve what we have not collected. We must enable access, the repository as persistence over time, we must secure the content, the archive as curation, and we must take care of the content, the repository as steward. Born digital content comes in an ever-expanding array of forms and formats. Consider just the following examples, and
this is exhausting: Published and licensed works such as e-journals, e-books, e-videos, and e-audio from commercial and trade sources, from academic publishers, from the growing array of independent and small publishers and distributors and the revolution in self-publishing and self-distribution. Further, the output of e-government, online learning and training materials, research data from universities and corporations, social media and all of its wonderful expressions, electronic archives that come with personal papers and organizational records, including e-mails and manuscripts and business papers and financial information, websites and web documents, visual images, spatial data, longitudinal observations, software applications, both proprietary and open source, video games, medical data, with the inherited challenges of patient privacy, live feeds like RSS and news information from around the world, visualizations and simulations, interoperable metadata like MARC and BIBFRAME and schema.org and so on and so on and so on with so many new things that will grow in intensity and intricacy. The people who look to us to capture, organize, and preserve stuff are going to be really pissed because we have it taken control collectively to solve this problem.

Commandment number two: Thou shall fight the information policy wars. We, libraries and publishers, must more rigorously represent and advance the public interest and needs of our users and readers in critical information policy areas. In national and global forums, we must embrace an expanded role in the legislative, legal, and political arenas, but too often I think we find ourselves in conflict with each other. Quentin Crisp, the British eccentric, was giving a talk in Northern Ireland, and he mentioned over the course of his talk that he was an atheist, and a woman popped up during the question period and said, “Mr. Crisp, can you tell me whether it is the God of the Protestants or the God of the Catholics in which you don’t believe?” We need to have our act together. The policy areas of interest are numerous and complex and include intellectual freedom and concerns over censorship, privacy and civil liberties, government financial support for education and research programs, including library funding, access to government information, network neutrality and telecommunications policy, open access to research and educational context, and copyright and intellectual property. This is, as one presidential candidate might say, “Huuuge!” Copyright is a topic of particular concern. Broad exemptions for libraries like fair use, though strengthened by recent court decisions, in particular limitations of the law which allow us to do such things as make copies for users and interlibrary loan and access for the print disabled and preservation, they’re all under threat. There is increasing focus on international agreements and treaties that influence our national laws and may not support our historical principles. More and more of the publications and databases being provided by libraries are covered by the private law of contract and not by the public law of copyright. Technological controls and digital rights management systems are reducing the ability to apply fair use and other valuable exceptions in the law. How can we play a substantive political role in these information policy areas?

Three: Thou shalt be supportive of the needs of our users and our readers. We, libraries and publishers, are developing a more sustained and actionable understanding of our user communities. Who are our users? Probably much more diverse than we realize. Where do we intersect with our users? Way beyond the walls of our physical spaces. How do we know about our users? Current tools of measuring and surveying and observing and listening are probably inadequate. As the late newscaster Charles Kuralt once noted, “Thanks to the interstate highway system in the United States, I’m able to travel from New York to San Francisco and see absolutely nothing.” The infrastructure is important but totally inadequate. Users want more and better content, but they want more and better access. They want convenience. They what new capabilities, the ability to manage costs, participation, and control over their own information environments and individual and organizational productivity. Users want technology and content ubiquity. They want web-based services with no lines and no limits to service. They want technological sandboxes, places for experimentation and fun, but also privacy spaces, places with protection and anonymity. They want support services, help when needed at appropriate levels of expertise. They want guidance to community resources and assistance with health issues and jobs and careers. Our users want us in the library to be authoritative and expert, trusted sources. They want us to be authenticated and secure, appropriate and pertinent, that is reputable and relevant. They want us to be accessible and omnipresent, that is always there, everywhere, when they need us. They want us to be at advocative, that is supportive of the diversity of
needs but also a voice of shared interest. They want us to be audacious and attentive, that is bold and innovative but not way out in front or too far behind where they are. How can we help users attain their goals, achieve well-being, realize benefits, move forward, make personal connections, participate fully, and have significant effect in the world through us, through libraries, through the content and tools that we work with publishers to provide to our communities? We don’t draw a line between what we do in libraries and these truly human requirements and expectations.

**Fourth: Thou shalt cooperate in more rigorous ways.** Cooperation is part of the professional DNA of libraries, but we need to move from “Kumbaya” to a much more radical strategy for collaboration. We know how to collaborate on a significant scale in such areas as cataloging and interlibrary loan and document delivery and licensing of databases, for example, but we need a deeper integration of operations in areas of mass production where we have hopeless redundancy across the library community and early co-investment in new infrastructures and new initiatives, not building it only at the institutional level, and in a commitment to a shared network, a shared complementary network of centers of excellence. From the conditions of knowledge scarcity over the centuries to the oppression of knowledge overabundance in today’s and tomorrow’s library, cooperation has been and will become a much more constant for service, success and survival. The future health of the library will be increasingly defined not by sharing resources on the margin but by new and energetic relationships and combinations and in innovative entrepreneurial partnerships. The measure of success of collaboration must be quality, productivity, leadership, and transformation. We are at a period of what I call polygamy; of rampant partnering and combinations. We’re marrying everybody in order to get the job done. We’re moving into a period of what I call “parabiosis.” Think of Siamese twins sharing body organs, body parts, body systems. Deep pairings of libraries and their resources. We’re advancing toward a period of what I call “particularism,” with powerful disciplinary service technology and workflow specializations across the library community. Let’s remember that every snowflake in an avalanche pleads “not guilty.” This is a shared responsibility.

Several years ago, I published a paper entitled “Symbiosis or Alienation: Advancing the University Press Research Library Relationship through Electronic Scholarly Communication.” I believe the evidence, ideas, and strategies outlined in that article can be exported to the current and future working relationship among all types of publishers and all types of libraries. I called at that time for a shared plan for collaboration, joint publishing initiatives, shared information policy agenda and coordinated advocacy work, joint consultations with researchers and authors, continuing education and training programs, content licensing principles, technology and metadata standards, usability testing, research and development projects through grant funding, preservation and archiving cooperatives, the management of born digital scholarly works like research data. But we have made very little progress in building this partnership between the library and the publisher community. And it’s almost like what the Episcopal Bishop said to the Baptist Minister, “Brother, we both serve the Lord. You in your way and I in His.”

**The fifth and final lost commandment: Thou shalt work together to improve knowledge creation, evaluation, distribution, use, and preservation.** I don’t know how He got that one on that tablet, but it was pretty long, right? For this commandment, I’m going to briefly focus on the scholarly communication process and the working relationship among researcher, publisher, and library. Researchers have the urge to share the results of their research through publication. This is the way they communicate with scholars and students around the world. It is part of the academic culture in which they have been raised. It is the way in which their ideas and contributions are preserved for future generations. It is their source of prestige, recognition, and remuneration. Researchers are telling us they need support in several critical areas. They are seeking assistance in navigating, analyzing, and synthesizing a literature they simply cannot keep up with, especially when they move into new and multidisciplinary fields. They want guidance on working in an open research environment with scholarly exchange that is continuous. They require more robust expertise databases, subject ontology’s and researcher information systems. They expect more consultation and support with research data management, which they know is increasingly mandated. They want help with awareness and
integration of disparate sources and gray literature. They argue for an informationalist and partner model for library support. The library community has been standing on the side of the scholarly communication stage for decades. We’ve been screaming like a Greek chorus, “It costs too much! It takes too long! We give too much away, and nobody is listening to us!” But this is an important public policy issue. Scholarly communication embraces communities of creation, production, distribution, consumption, and use. The publisher community has largely controlled the production and distribution channels, while universities have funded the creation, consumption, and use. Libraries have long argued that we are choking on the proliferation, that we need to rethink the location of the quality marking, that a corporate economy has consumed what was a guild economy, that scholarly publishing is largely a dysfunctional monopolistic market, and that we are not able to advance new models of digital scholarship. We have bemoaned what Larry Lessig has described as the constraints on access to information: The market, that is the cost; the law, that is copyright ownership; the technology and the norms, the way we have always done it. We bring, I think, some sustained core interests, a more competitive market to reduce costs and increase innovation, easy distribution and reuse of publications for purposes of scholarship and learning, innovative applications of technology, quality assurance and integrity, and permanent archiving of the scholarly record. Open access flows out of the 1990s in the library community. I remember vividly at an Aero membership meeting Cliff Lynch throwing open the proverbial window and telling us, the directors in the room, that if we were mad as Hell, than we had better do something about it. And from there the philosophy, the strategy, and the practice of openness in the research library community was born, and the SPARC organization was launched. We now talk about open scholarship, open data, open source, open educational resources, and so on, but in spite of significant investment and federal mandates and new publisher policies and researcher commitment, open access remains very much a work in progress.

So where does this extended commentary take us? For me, it means that over the next decade we must forge a new economy for libraries and publishers. I went to a play recently in Manhattan called “Extinction.” And there I learned that there are actually two types of extinction that biologists talk about. First, there’s terminal extinction where the species disappears. Second, there is phylectic extinction—who knew? Where a new species will evolve. I would argue we, me, must be committed to that phylectic extinction goal. Libraries must invest more resources in space, in innovative technologies, and in staff expertise while also assuming new responsibilities for such things as digital preservation, new services to scholars, research and development, and deeper involvement in learning. We also need to understand how to think about playing a larger role in the discovery space, and Anja (Anja Smit, prior plenary speaker) talked about that. We need to question: Should we? Do we want to? Our users will evaluate us not on the things that we can provide but by the things we can enable them to do and accomplish. Libraries increasingly do not and will not pay for content that is simply not used: Content and articles that are not read or cited and books that are super specialized and have narrowed scholarly interest. This is a flawed and unsustainable market. Go directly to the students and faculty and see if they’re willing to purchase or pay-per-use even with university subsidies. Open content will be more available and accessible. It will not, in my view, displace commercial and scholarly publishers, but more and more higher education institutions and funding organizations by policy and by law will mandate openness. Digitized historical databases will be delivered through national and global platforms unless publishers can provide significant added value.

Albert Einstein, when he came to the United States, would never fly in a plane. That scares me that a physicist would not fly in a plane. So, he went often by train. On one such trip, the young attendant was coming around to collect the tickets. He came up to the seat where Einstein was sitting, and Einstein started to dig in his pockets and look for his ticket, and he said, “Oh, no!” He recognized who this guy was. “Dr. Einstein, please, you don’t need to find your ticket. Don’t worry about it. Just don’t worry about it.” But, Einstein persisted. He crawled on the floor. He lifted up the seat, and again the young man says, “Dr. Einstein, come on now. You don’t need to find your ticket.” Einstein whirls around, looks him in the eye and says, “Young man, it is no longer a matter of whether I can find my ticket. It is a matter of where I am going.” So, my final point is we need to spend far less time looking for our tickets and spend a lot more time thinking about where we’re going. Thank you very much.
The Devil Is in the Details: Challenges of Collaborative Collecting

Judith Russell, Dean of University Libraries, University of Florida

The following is a transcription of a live presentation at the 2016 Charleston Conference.

Judith Russell: Thank you for having me here to talk with you this afternoon about the challenges of collaborative collaboration and also the rewards because there are significant rewards, and I hope you will see that as we go forward. There is an old African saying right here on this slide: “If you want to go fast, go alone. If you want to go far, go together!” This saying is mounted on the front of my computer monitor thanks to a fortune cookie that I ate many years ago, and it is there for a reason. It reminds me of the benefits of collaboration and also that I need to be patient with the process. Those of you who know me or know of me will recognize that patience is not the first characteristic that comes to mind when my name is mentioned. My instinct is to want to get things done quickly, and I do need a reminder that patience is a virtue.

The libraries at the University of Florida are very active participants in a number of collaborative collection development initiatives, and they do provide significant benefits to us, to our partners, and to others who can benefit from the information that we gather. I’m only going to touch on a few of them today, but I’ve tried to pick ones that represent kind of a range of things. I think it’s really important, though, to start out by recognizing that each of them is requiring this effort to establish and sustain trust and to maintain the value for the other collaborators, not to drift away, not to lose sight of the fact that everyone has to be benefiting, and also that it will take longer. It always takes longer than I want it to and always longer then I think it should, but that is because we do have to consult a number of people, and we have to maintain our awareness and concern of their preferences. But that has not stopped us from actively seeking to identify and participate in these initiatives, but again, we have to remember that everybody has to benefit, so there is a need to constantly revisit and think about that.

This summarizes our library mission and vision, and I wanted to speak about just the top two bullets. One of them is the second bullet that says, “We initiate and participate in collaboration and community building.” This is in our DNA. I think it is in the DNA of most libraries and most librarians, but it is very much in our DNA, and it does drive our openness to these collaborations, and I think helps to make those collaborations successful. And then the second one is this issue of offering key services at the point of need, and increasingly, that means dealing with digital content so that we can have the content available for access anywhere and anytime. So, we do favor electronic content. We dedicate a huge percentage of our materials budget to electronic resources. We have our own digital platform, the UF Digital Collections, which already has over 12 million pages of content and is adding about a million pages a year. So, we do take this seriously as the digital content being an important part of who we are and what we do and how we contribute.

We do a lot of collaborative acquisition, as again I’m sure many of you do. Over 50% of our material’s budget is used for collaborative acquisitions. We’re doing a lot of patron-driven acquisitions in bilateral and multilateral arrangements. That does mean we’re focusing less on the future needs of our researchers and more on the needs of current users, sort of the nature of the beast. We do also have a shared service viewer that runs our integrated library system and provides other resources, so it facilitates sharing of particularly print resources among the academic libraries in Florida. And we do have other consortial relationships like HATHI Trust. So, we do look at collaborative acquisitions in a lot of different ways.

But, today I really want to talk about other types of collection development initiatives, and I picked six of them, well six-ish; some of them go a little broader than the first bullet. These are examples of several major collection development initiatives, each of which contributes to meeting our institutional needs and those of our partners. They do benefit all the participants; the benefits may not be equal to each participant, but there still has to be a benefit to each of them to engage them and keep them in the project. The first and third, the digital library of the Caribbean and the ASERL Collaborative Federal Depository Program, had already begun when I arrived at UF in May of 2007, but both have
expanded considerably since that time, and as often happens with good collaborations, dLOC has expanded to additional and important projects that can be reviewed independently and cross searched within dLOC, as you’ll see as we go forward.

dLOC is now 11 years old. It’s a long-standing collaboration of libraries, archives, and research institutions in the U.S. and the Caribbean and even a few in Europe. It reflects the diversity of the Caribbean, and it supports a wide range of content in many languages. One of the major challenges to dLOC, other than those identified in the bullets, was the abrupt transition from grant funding to self-funding when the final two years of the grant were canceled in the recession. We made that transition successfully in part because of the willingness of the partners with greater financial resources to establish a tiered due structure, but that needed to be done with great care to respect the feelings of all the partners and avoid any implication of lower dues indicating less importance or status, keeping the board as inclusive as possible, and having the leadership include well-respected Caribbean partners was, and continues to be, essential. This is the homepage for dLOC (http://www.dloc.com/). I would encourage you to go and visit it. It shows here some of the sample of collections, two that I would call particular attention to are the one at the bottom of the page, the “Voodoo Archive,” which is a very interesting digital scholarship project that was developed by a faculty member at the University of Florida, and another that doesn’t show on the screen but which is called “Haiti and Island Luminous.” It has received numerous awards and is quite interesting and diverse and really worth looking at.

So, this, dLOC, is a perfect example, looking now at the Haitian law and legal materials, of the serendipity that occurs often in collection development in identifying opportunities for collaboration. Several years ago, I was having lunch with Jerry DuPont, who is the founder of LLMC, and he was late for lunch because he was at a planning meeting for the Haitian Law Initiative. I asked why he hadn’t contacted UF, and he said it was because our law library did not collect heavily Caribbean law, and, of course, LLMC was a consortia of law libraries. I responded that the Latin American and Caribbean collection in my libraries did collect legal materials. LLMC was also looking for a public access platform so the people of Haiti could access the materials that were being digitized for their benefit. Their platform at that time at LLMC was only open to members. dLOC became the public access site, and LACC, our collection, became the third largest source of content for this Haitian law project.

Out of that project grew another project modeled on that which is on Cuban law, and that one began with my collection, so we were the first collection that they mined, and we are now expanding it to other partners. So, again, an example of how one collaboration leads to another. And in turn those collaborations related to other things we were doing and have resulted recently in the establishment of the Cuban Heritage Digitization Project. Over three years ago, on his first visit to the Smathers Libraries, Eduardo Torres Cuevas, who is the director of the Biblioteca Nacional Jose Marti, the national library of Cuba, signed an agreement to join dLOC, and we began to exchange digital files and host his digital files and to plan for collaborative digital initiatives. We already had digitized a number of Cuban newspapers that he didn’t have and gladly provided him with digital copies. He has a collection of Cuban American newspapers from 1890 to 1930. There are only five or six issues of those newspapers held anywhere in Florida, and he has a very good collection of them, so he is digitizing those to give to us, so it is sort of ironic that we’re giving him Cuban newspapers and he’s giving us Cuban American ones, but it works. Right? So, out of several years of collaboration and exchange visits came a recent Convenio Contract to create the Cuban Heritage Digital Collection. BNJM has digitized 89,000 cataloging records from 1900 and earlier, and 58% of those records, according to their analysis, are for materials uniquely held in their library. He has committed to digitizing his unique holdings and providing them for public access through dLOC. He will also host them locally, but Internet access in Cuba is extremely limited still, so he doesn’t feel that he can be the platform for worldwide access and not even always for the best access for Cubans. We’ve agreed to collaborate with other libraries to digitize as much of the remaining 42% as we can, adding them to the collections of dLOC and giving him digital copies for local use.

I want to switch now to another collaborative initiative that was underway when I came to Florida, the ASERL Collaborative Federal Depository Program. I immediately joined the committee that was
managing the initiative under an IMLS grant. It was a demonstration project to consider ways to improve collaboration among documents librarians in the region. I was particularly well-suited to participate and then provide leadership for this project, as my immediate prior position had been as the superintendent of documents at GPO. These three bullets are key elements of the program: Stay within the law but not necessarily be constrained by tradition or GPO policy that was not legally mandated; seek to build retrospective collections only for what we called “centers of excellence.” There had been a value system in the depository program that said that regionals should try to build comprehensive retrospective collections so that all of us who are regionals should have every document we could possibly get, and that was just overly burdensome, duplicative, and inappropriate. We also wanted to simplify and harmonize disposition rules to reduce the burdens and facilitate the transfer of needed titles among libraries throughout the regions. The tradition had been that each regional set the rules for the libraries that were selective within their space, and so I think there were 49 regionals at the time, which meant that there were 49 different sets of rules. So, for us, there were 12 regionals in our 10 states, which meant 12 sets of rules. We had some struggles with GPO to get the program approved, but it has been very successful.

We started on this initiative, and then following the release of an ITHACA report on the FDLP in October of 2009, there was a discussion among the ASERL Deans that lead to a decision to build on the IMLS funded project and develop a regional initiative that pushed the boundaries of the FDLP enabling a legislative but remain compliant with the law. So, we wanted to go to the edge of the cliff, but we did not want to jump off or push anyone off. It certainly helped that I was very knowledgeable about Title 44 and the Depository Program and had been working for many years on ways to improve the program without the necessity of statutory changes. We expanded the governance to include documents librarians from all 12 of our regionals, even though two of them were not ASERL members, and also we included other government documents librarians from selective depositories because we needed to understand and meet their needs, and we set up a steering committee of deans from both regional and selective depositories. This insured commitment both in terms of resources and permission, in fact encouragement, to act boldly and it also provided a means to settle questions that the documents librarians could not resolve among themselves. The principal way that this worked was to set up Centers of Excellence, and the idea of a Center of Excellence was that we would take out parts of the collection that were relevant to the university and to our constituents, and we would build out those collections. We would invest our time and energy in cataloging and digitization and developing reference skills and so forth around those collections. So, I’ll give you two examples: Ole Miss has a major archive for the Institute of Certified Public Accountants. They chose the IRS. They felt that focusing on documents published by the IRS would enrich and complement the ICPA collection and vice versa, so it marched with the needs of their university and the interests already there. Florida International University was already operating the Everglades Digital Library, and they decided that they would take a subject focus, and they would look for any document by any federal agency that dealt with the Everglades. So, they weren’t focusing on an agency. They were focusing on a topic.

An important part of the project is the commitment to fill identified gaps in the CLE collections. That is the one place where we were committed to retrospective digitization, and one of the biggest burdens on depository libraries is that legislative mandate to first offer publications to other depositories before you can weed or discard it from your collection. So, to address both of these requirements, UF developed and continues to host the ASERL Disposition Database. Because there are common rules for disposition, the process is automated, and it matches offers of materials from libraries that are planning to discard with needs from libraries that are trying to build collections, whether they are CLEs or not, but preferential access to those discards is given to filling gaps in CLE collections. In 2015, 312,000 documents, or groups of documents in some cases, were offered. 21,000 were claimed, of which about 15,000 were claimed by Centers of Excellence. So, that may seem like a small amount that we only saved 21,000 documents from being discarded, but that is way more than had been saved under the old manual process, and they were going places where there was a high need, and there is a lot of duplication in those collections, so it is not surprising that there would be a lot of discards as well. Many more items have been adopted under this program than the old methods, and there is still a very heavy volume of discards.
So, also to fulfill the IMLS Grant, we shared the cost with ASERL to develop another piece of software tool called the “Gap Analysis Software,” and it lets us compare records from various libraries to identify unique holdings. This does facilitate identifying gaps because if we see that another library has a publication from one of our Centers of Excellence collections, and we hadn’t already identified it as missing, we can add it to our needs list, and we can consider, if they are not able or willing to give it to us, at least settling for a print or digital surrogate so that we have a copy in our collection until we can get an original.

So, we are, as I said earlier, a regional depository. We serve Florida, Puerto Rico, and the Virgin Islands, which crosses back over into our Caribbean interests. We are considered a multistate regional because we are serving institutions from more than one geographic area. We are very active in the Collaborative Federal Depository Program. We have 35 Centers of Excellence, which you can see listed here. We are a land-grant institution, so not surprising, a lot of interest in USDA agencies. I worked at one time in my career at NCLIS and OTA, so we’ve adopted them. Our biggest commitment has been to congressional hearings, and we are still working to get those catalogued. That is an enormous collection, but the Panama Canal was particularly interesting. We have a very significant collection on the Panama Canal, and we merged into our collection a collection from a small museum on the Panama Canal, and so we are taking care of the Panama Canal Commission and its predecessor agencies but also, not unlike the Everglades example, all other federal documents and maps about Panama and the canal regardless of the agency. It could be the Corps of Engineers. It could be a treaty from the State Department. It could be hearings. It could be the GAL report. If it is about Panama, we want to have in our collection.

We are actively cataloging and digitizing and making these things available for public access. We, in fact, have submitted to GPO the digital copies of all of our Panama Canal documents, and they are now appearing in FDsys (govinfo) depending which name you know it by. We do harvest digital or digitize content for our Centers of Excellence, but we don’t otherwise harvest or host digital content locally, and that is really true across our collections. If we send 100 brittle books out to be digitized, and 10 of them were from our preeminent collections, we would bring those back and host them locally, but we would be perfectly happy if we’ve used Internet Archive, for example, for digitizing, to just leave the others at Internet Archive and link to them. We don’t feel compelled to host something locally unless it is part of one of our important collections, and we follow that same policy with the documents.

So, one of the commitments that I made as a Dean, which I some mornings wake up and wonder why, but we had 300,000 government documents uncatalogued in storage when I came, and we started cataloging them, and then in 2014, to make space for student services, we moved the remainder of the documents collection to off-site storage and committed to cataloging the entire collection. We have catalogued over 560,000 volumes, and we have created almost 14,000 original cataloging records which we have put in OCLC, so they are available to other people who are doing cataloging. Most of these records have been copy cataloguing. There is a lot of cataloged documents out there; it is just people have not applied those records to their catalogued materials. Based on the large number of original cataloging, people often comment that we have a high number of unique documents, and I remind them that I don’t necessarily have unique documents; I have uniquely catalogued documents because so many of us still have a large volume of pre-1976 documents that are uncatalogued. It is difficult to estimate how far along we are because the pre-1976 publications are interfiled with later ones, but we think we are about at the halfway point, so we still have a lot of cataloging to do. I did realize, though, that this was one of the best things I could do for my selectors and for the FDLP as a whole. By cataloging my collection, each of them can make informed decisions about their own collections. There was and is a large volume of weeding going on based mostly on the assumption that the regional has everything, or at least some other depository will have a copy of everything that is being discarded, and it’s an assumption without facts to base it.

Last year GPO started discussing the option to allow digital substitution for regional collections. Right now, regionals are required to keep everything tangible all print or microfilm forever. We stepped up and offered to serve as a preservation collection for print versions of digital or digitized documents in govinfo. When at least four preservation copies have been identified, other regional depositories will be
able to request authorization to discard the print and rely on the electronic. The reason we stepped up so quickly is that this requires minimal changes to our standard procedures. The biggest change is that the preservation copy in our system will become a noncirculating copy so that we will rely on the digital copies, and that is feasible because there will be digital copies. I think that is another important factor about collaboration. The more minimal the disruption is, the easier it is to get the collaboration to work and to get people motivated to participate in it.

We also created a “Last Copy” policy specific to federal documents. As far as I know, it is the only one in the country, but we are hoping that other people with either shared storage programs that include documents or other regions will adopt this policy. We feel that if one of our selectives or any of the academic libraries in Florida feels that they have a document that they can no longer retain but that it’s important that they still have access they should be able to send it to us, even if it is not within the framework of our Center of Excellence collections, and we should receive it and host it as a service because we are meant to serve our community. Because it is not fully catalogued, it is not easy for them to know if we have it, so they can’t determine whether we have it. They can send us a list, we will walk our shelves, we’ll verify what we have, and we’ll let them know what we don’t have, and they can send it to us. That has worked very well. We’ve gotten not a large number of documents, but we’ve gotten some, and it does build a lot of goodwill for the overall collaboration.

So, I think you can begin to see how these things began to link together. Participation in the ASERL Collaborative Federal Depository Program led to the transfer of our regional collections to what we call FLARE, the Florida Academic Repository, and the existence of FLARE gave us a logical home for the regional collection. So, we’re now managing government documents like we manage other shared monographs and journals. We plan from the beginning on the assumption that the collection would eventually service the private universities and the public and private colleges in Florida, not just the university system, and we wanted to make sure that our policies and procedures anticipated those future participants. So again, an important part of collaboration is a broad engagement and participation of future partners. We didn’t want to get all the rules developed so that we are perfectly happy and then say, “Hey, guys. Now we’re ready for you to join,” and have them say “Oh, but we can’t do it because this rule or that rule.” It is much better to have representatives at the table with us as we plan the basic rules and just put the right rules in place from the beginning. The FLARE collection is eventually supposed to be housed in a high-density storage facility in Gainesville. The planned facility will eventually initially have two modules with a capacity of 5.2 million volumes. The land adjacent to the current facility has certainly enough space there to build multiple additional modules with a capacity of over 20 million volumes as they are needed. I actually proposed building this facility to the Provost when I interviewed for the job at Florida, whether or not she chose to hire me. I had my first meeting with my new colleagues, the library deans in the state university system, in June and proposed it to them. We submitted a proposal to the Board of Governors in August, and it was approved enthusiastically in October of 2007. The land speed record for getting anything approved through the Florida Board of Governors, I might say. However, although we were given planning money in 2009, and then we had a recession, so this is perhaps one of the greatest tests of my patience because we are still waiting for funding. But, we have continued to build the FLARE collection using the two off-site storage facilities, and I’m optimistic that this might finally be the year when we actually get funded.

As noted here, the FLARE repository already has over 2.2 million volumes, and 1.2 million have already been trade and inventory. They’re already identified in OCLC and are available for interlibrary loan. Our policy is that we loan from this collection the way that we would loan from any collection in our academic libraries, so it is not a closed collection just to serve the participants. It is an open collection to serve the broader community.

Simultaneously with our development of FLARE, ASERL was developing its own print journal archiving, and we aligned the policies for that so that FLARE could be an active participant in ASERL. So, every journal that is in FLARE is also in the ASERL Collaborative Print Journal Archive. And then in 2013, the ASERL project merged with the WRLC Collection to create something that we call “Scholars Trust.” Out of self-interest, but also interest in my community and to avoid duplicative entry into databases for two journal archiving programs,
offered to develop software that would serve both originally ASERL then into Scholars Trust and FLARE, and it is called JRNL, Journal Retention and Needs Listing. The lessons learned from the dispositions database applied here that we were going to have journal collections with missing volumes, and we should be able to communicate with our partners so that if they were weeding, they can identify that they had a volume that we were missing, or that we had a volume that they were missing, and we could fill in the gaps in our collections. As of June, there were over 17,000 unique ISSN’s recorded in JRNL, of which over half were from FLARE, so FLARE is a major contributor to this larger collaborative program.

We have another special project that I initiated, which is sort of nested into this print journal archiving. There are 40 ASERL libraries and nine WRLC’s, so 49 libraries in this partnership. Only 10 of us are land-grant institutions, so we identified 1,000 journals in agriculture and related fields. We went to the 10 land-grant universities in ASERL and to the National Agricultural Library, and we are collaborating to establish archiving commitments for those agricultural journals. We’re excepting commitments from others, but it seems so much easier, again scale is an issue, so it is a lot easier to start this project with 10 libraries than it would’ve been to start it with 49, so once the 10 got it well under way, we are accepting commitments for titles from others, but the starting point was the smaller group. We already have 762 journals that have been entered into the database for retention commitments, so it is working well.

We do expect to grow the journal program by subject. We’ve already got another small group, I think its nine libraries, who are working on architectural collections. There really is an advantage to finding a scale. There is a lot more self-interest in having architectural librarians from nine of our universities collaborate than there is to go out to 40 of them and just say, “Tell me any journal you want to keep,” regardless of the subject. It’s almost too big to get engaged in, so that’s been very helpful to look at it that way.

The final example that I wanted to share with you is a very different type of collaboration, and it is between the Smathers Libraries and commercial publishers, making it a little bit more controversial, maybe a lot more controversial to some. It began with a collaboration with Elsevier, which has expanded to include other publishers through CHORUS, and there is a fairly detailed article about the Elsevier project in Collaborative Librarianship in Volume 8 published in early summer of 2016, so you might want to look at that if you want more details about this project, but kind of here is the who, what, why, when, where. So, it was a bilateral project with UF and Elsevier. We were looking at each of us having goals which were at least not incompatible. They were sometimes different but not incompatible, and the ultimate way that we did this was to link articles and download metadata into our institutional repository using a free API, application programming interface, from Elsevier. So, we did have some common goals. The way this came about was in conversations with my Provost, my Vice President for Research and the Faculty Center Research Council over several years. I was asked several times why the universities couldn’t solve this problem, the problem being identifying UF faculty research publications with minimal burdens on the faculty. Don’t ask the faculty to tell me what they published. Don’t ask them to send a copy of the article or the manuscript to me; find a better way. There are many more academic faculty than library faculty. I assume that happens to you as well, and they are very productive. Our faculty produce over 8,000 journal articles a year, so we also need a solution that place minimal burdens on the library faculty and staff.

Since the UF authors published between 1,100 and 1,300 articles a year in Elsevier journals, I first approached Elsevier to see if I could obtain author manuscripts directly from them. Think of it. They have the manuscripts, right? I don’t have to go to all these different authors. They know what permissions were or weren’t granted, and it seemed very logical. I’ve quickly learned that Elsevier has recently developed APIs to facilitate identification and downloading of metadata into local institutional repositories and was looking for a partner to test them. My staff quickly decided to use the APIs rather than to seek copies of the manuscripts, and I have to say UF did not have a culture of deposit. At the time that we started this, there were only seven Elsevier articles in the institutional repository, so out of an average of 1,100 to 1,300 a year that were being produced, our goals in phase 1 were to increase the comprehensiveness of our coverage of Elsevier published content in the IR, to provide subscribers with access to what we termed the best available, that is the published version of the article. Our faculty definitely preferred that we take
people to the published version of the article with all the links and other things that were embedded in it. We wanted to integrate the published articles into the IR at UF so that people could find other content that was already in the IR, whether it was a conference paper or something else, and use all of the content that we had on that topic, not just the content that wasn’t published in commercial journals or just the commercial journal content. So, in phase 2, which we just started recently, we’re now providing an access option for users who do not have a subscription to Elsevier. We’re providing an option for nonsubscribing users to see a manuscript version of the article, and we are also doing full text indexing, wherein the first phase was indexing just with metadata and abstracts. And we’ve been doing usability testing of the first phase, and so we’re looking to that to help us change and improve the interface, and we’ve been doing research on open access publishing by UF authors and using Elsevier metadata for other university purposes, including compliance.

Right now, we have over 30,000 articles by UF authors from 1949 forward that we were able to download from Elsevier. Again, metadata and abstracts, linking to Science Direct, full text access, full access for users with subscriptions, and we are working on phase 2 of alternative access for other users. Ninety-five percent of the attempts to use Science Direct come from subscribers. There is only a 5% rejection rate, so there really was a sense that this historical number of denials for nonsubscribers was relatively small, and, therefore, we didn’t really deal with that in phase 1. Often that denial is corrected by logging into a VPN server. People are sitting at their desks, getting right through. They go home, they sit at the desk, and they don’t get in, and when they get rejected actually says, “You might consider signing on through VPN.” Often when they do, they then get in, but we are now focused on serving those nonsubscribers.

The benefits of collaboration were a little different for each of us but one that both of us were very interested in was how this might ultimately facilitate oversight of and participation and compliance with the new public access mandates that were coming up. We also recognized that the university was going to need help with this compliance, and so we were testing how having this metadata might help us at least inform them of what articles UF authors have published so they would note it and try to see if there’s been compliance with those articles. Our phase 1 goals were, as I say, focused on metadata and indexing of metadata. The most difficult problem we encountered was identification of UF authors. I’ve always been a good strong believer in ORCID. I’ve become a good strong believer in Ringgold. You know, we’re not at Clemson or Emory or Stanford or someplace that has a unique name. There is a number of universities that have the word “University” and “Florida” in their titles, so it’s a big effort on our part and Elsevier’s to get rid of the ones that were University of Central Florida or Florida State University or the false traps. We think we’ve pretty well overcome that, but obviously, better metadata that identify both the authors uniquely and the institutions uniquely would be enormously helpful.

The surprising discovery in all this was the high number of open access articles published by UF authors in Elsevier journals. We’re doing further analysis on the open access publishing, and we expect to survey these authors to determine more about their motivation and their source of funds. We hope some of them will become champions for our campus-wide open access policy that we’ve been trying to get through for several years. Interestingly, out of these 601 articles by 1,443 unique UF authors, so that is not counting their collaborators from other institutions, which are in some cases also participants, 8 to 10 of these had more than 20 open access articles in Elsevier journals in the years between 2009 and the early first quarter of 2016. That seemed amazing to me. One of them had over 30, so I really want to meet these people. I’m very curious about what they’re doing. Obviously somebody has got a real strong commitment.

So, I wanted to show you this because one of the issues was how would we display this content, and the Elsevier people wanted to be sure that people knew when they had access. Actually, if you do a Google Scholar search, it does not tell you if you have access. If you click through, it will stop you, but it doesn’t tell you before, but Elsevier felt, and we agreed, that it will be helpful if somebody knew. It could be very frustrating if you click through a lot of things but kept getting told, “No, no, no.” So, we came up with this symbol that you see at the upper part of the screen. It says “Publisher version. You have access.” And then we also wanted to let people know when they might not have access, but because it is sometimes issues like VPN, we didn’t want to say, “You don’t have access,” and maybe stop them from even looking, so we came up with a publisher version.
check access as shown at the one at the bottom, and you may not be able to read it from the back, but that article was published in 1990. UF does not own the back file for that journal. Consequently, the Elsevier API identifies correctly the user accessing the article from the UF IP range may not be entitled to access. The user from another institution that does have access to that back file would’ve seen the message “You have access,” so the API presents the results that are specific to the status of the individual user, which we all thought was really very beneficial, but it wasn’t kind of a guess. It was a very real and targeted response. And although, as I just mentioned, we have a lot of open access articles, we didn’t think it was enough to just say “You have access.” We really wanted to call attention to the open access articles, and so there is a specific and different label for articles that have open access, which obviously everybody can have access. We’re now working with Elsevier on the tagging in phase 2 for how we’re going to inform the users without subscriptions that they have access to the manuscripts, so that is kind of one of our phase 2 activities.

So, moving kind of the transition from Elsevier to CHORUS, when I first saw this slide, the word compliance was as it is now in the center, but there was one less figure on the slide, and I went up to Howard Ratner, who had used this in a presentation about CHORUS at an SSP meeting, and told him that the person on my campus who was losing sleep over compliance wasn’t on his slide, the Vice President of Research. I acknowledged the need to initially focus the development of CHORUS on publishers and funding agencies but pointed out that this was, in fact, a three-legged stool, and that there were three figures from academic institutions who represented one leg of that stool, the librarian, the VP research, wants a place there and the researcher, and they were not participating in the design of CHORUS. So, I suggested that when they were ready, UF and other academic institutions should participate in the development of CHORUS to ensure that it met our needs as well as those of publishers and funders. A few months ago, Howard called me and said they were ready to expand the partnership and asked if UF could participate, and I quickly agreed.

This is another example where the Elsevier project led to the CHORUS project. Elsevier is a CHORUS board member, and their member of the board kept them informed about our pilot, and, as the other publishers learned about how the Elsevier pilot was working, they became more and more confident about doing something similar with the universities through CHORUS. So, we now have a whole group of seven CHORUS members who are participating in the pilot, and we expect that more will participate before the pilot is concluded. There are others who are actively considering.

We have a variety of aims for the project, but from my point of view, the most important goal here is the goal of facilitating compliance. So, the way this is going to work is that we will be using tools to identify UF authored articles. Then CHORUS will in an automated way check the metadata for the funding source. Once the funding source is identified, they will verify that there is deposit in an appropriate funder repository or not, and they report the data to UF through this dashboard, and we will then turn that information over to the Compliance Office and the Office of Research. So, for example, Dr. Smith has published an article that was funded by DOE. They find that article, and they confirm that the grant was a DOE grant, and they look in the DOE repository, and they say, “Oh, yes. That is there.” So, the report that we get is, “Here’s the article, here’s the funding agency, and it has been deposited.” Dr. Jones published in USDA, but the USDA grant is not yet in the repository. The Office of Compliance knows at least at some point they will need to follow up with Dr. Jones to say, “Remember you need to deposit that article.” But, it may be that in the next month’s report. We haven’t decided whether these reports are weekly or monthly, but in the next report, it may then say now Dr. Jones has deposited, so this will be really important. My Vice President of Research is enormously happy about this because he sees this as being a very efficient way to help his office of compliance manage the fact that we have these 8,000 journal articles being written, many of which will have to be deposited in one or more repositories under these agency mandates.

The most frequent question I get asked about these projects is the one on this slide. Why Elsevier? Why CHORUS? And I think this slide answers that question. Eight of the 10 publishers most selected by UF authors are CHORUS members. Elsevier has the largest volume of any publisher, so naturally I started with them. Also, they were ready and willing to test it. But, there’s also Springer Nature, who has a large volume. Wiley has a large volume. Automated solutions for identification of access to
articles by UF authors from multiple publishers reduces the burden on the academic faculty. It reduces the burden on the library faculty and staff, and by sharing the data on compliance, it reduces the burden on the staff in the Office of Compliance within the VP of Research. So, I think you can see here that it’s logical for us to think about how do we get automated tools that allow us to do this efficiently? This is going to be something that every one of our campuses is going to be having to deal with, and so testing these things and trying to perfect them and make them into standard tools that we can all use I feel is very important. And I think it is very important that we were in it at the beginning of the process so that they aren’t developing these tools in a vacuum. They are developing them, and we’re already giving them feedback about what do we need on the academic side, which is going to be very different than what the agencies need to be reported and what the publishers need for their own information. So, we’re looking at all of the users and what their needs are and trying to address them collaboratively. It’s just in the early days. We’ve only been at it for a couple of weeks, but it already holds great promise, and we’re really excited about it.

I want to go back and wrap up by my African saying: “If you want to go fast, go alone. If you want to go far, go together!” I hope these projects give you some good examples of the benefits of collaborative collection development as well as to specifics about some of the initiatives in our libraries and with our partners. We share expertise, we share costs, we challenge one another, and we stimulate creativity and innovation, and we do not only go farther, we go better by going together. It is not without cost, sometimes in dollars, more often in commitment from our library faculty, but it is absolutely essential. We could not do most of these things effectively on our own. They have to be done with collaborators to really benefit from them.

I’m back again to the other slide about libraries, and I here emphasize the bullets, whereas on the first one I was emphasizing the lead statement, but these initiatives do require effort. They’re not without pain. We have to sustain trust. But I’d like to talk for a minute about what makes a successful collaboration and get back to the title. The devil really is in the details. Ultimately, every collaboration is a risk. They’re not going to all work well. They won’t all be sustainable. But we have to be open to the potential and able to identify the ones that are likely to pay off and willing to recognize the value of the lessons learned from the ones that don’t. Sometimes we learn more from a failure than we learn from a success, and we have to be able to take that and learn from that and go back to the well and try again in a new way and see if we can go forward.

Each project requires leadership, vision, and commitment. We need to be able to see the potential, to identify the benefits in our own institution and to the others who seek to join us. We need to be able to convince others, both internal and external, to participate, and we need the patience to persevere even though it is likely to take longer than expected. As leaders, we need to provide the necessary resources and empower and encourage, and I’m guilty sometimes of badgering the participating staff. I will say that most of the time when I question the staff, I’m asking “What do you need from me in order to move forward faster? What do you need from me?” But it is still badgering, so I acknowledge that. I’m more patient than I was but maybe not patient enough yet. We do need dedicated believers and implementers to understand the vision and work out the details and strengthen the collaborations. They’re the ones who make it happen. I can have the vision. I can go to Howard Ratner and say, “Let me in.” He can say, “Come in.” But, it is only when his staff, and my staff, and the staff of these publishers sit down at the table and really talk it over and start working on it that it will actually happen.

We often face uneven resources, skills and commitment among our partners, and we need to accept that and adjust to it. International partnerships, and sometimes even domestic ones, can have language and cultural barriers. Working with the publishers certainly identify different vocabulary and different perspectives, but these were not barriers, and we both learned from exploring the differences and gaining greater understanding, and in the end, it helped us to move forward. Successful collaborations build on each other. dLOC into the Haitian law and to the Cuban law and to the Cuban heritage, FLARE into ASERL, print journal archiving and the Scholars Trust, Elsevier into CHORUS; so if you look at these things not just in their isolation of this project but where does this project take us? Also, a successful collaboration builds trust and makes the next collaboration with that partner or perhaps with observers easier. They
surprise and delight. I cannot tell you how enthusiastic my staff are about some of these projects. I mean, they just feel so good about what they’re getting done, and so even though it is hard work, the rewards both personal and professional are great. They meet new people. They form new friendships and relationships, so I can only say you should try it. You’re going to like it, and it is much more fun to travel with others than to travel alone, even though it will slow you down a little bit. And I will say to you also that if you have a great idea and you need a strong collaborator, I hope you’ll consider sharing your idea with us. We all have limited resources, so we can’t join everyone that interests us, but we do help and participate where we can. So, I hope we left little time for questions. Thank you.

Kalev Leetaru, Senior Fellow, Center for Cyber & Homeland Security, Georgetown University

This is a transcript of a live presentation at the 2016 Charleston Conference.

Kalev Leetaru: Thank you so much. It is a true honor to be here with you all today. Most of my career over the last 20 years has focused on how do we reimagine society through the massive amounts of data that we have with us? And I want to open with this project here.

This is a project I did several years ago. I approached the Internet Archive and said, “You know, I want to reimagine the book.” What if we thought of books not as containers of text but as the world’s greatest art gallery? Lots of people in the past have explored pulling images off of digitized books, but I wanted to see what would happen if we did that at scale. So, in the end, we took 600 million pages of books dating back 500 years pulled from over 1,000 libraries worldwide, and took the existing OCR that the Internet Archive had already done, and pulled off every image off of each page and all the text that surrounds that. What we can do is really fascinating things, finding fascinating examples like this, which came from, I forgot the era, but it was apparently a wedding (referencing image on slide). We can find fascinating things like to remember that 100 years ago, mail by train, taking your letter and sticking it in a train going by, was this incredible way of speeding delivery of messages compared obviously to 100 years prior. Being able to walk through 500 years’ worth of imagery, and the image in the top left there is a woodcut from 500 years ago, but the image to the right is a very famous one, the telephone and communication, and it’s really fascinating. One of my favorite books, 1921, there was an illustration of a teenager using a telephone, and the book said, “In the future, we predict that teenagers will actually make use of the telephone heavily.”

It’s very funny to really see it visually, to be able to see how things are depicted, and famous illustrations through time, but also images like this. So, these two images, these are from Emblem Books, and the two books I think are about 100 or maybe 200 years apart; one from a library here in the U.S. and one from a library in Europe. Now, historically a historian would likely never have encountered both of these two images from these two books from these two libraries. It’s unlikely that someone would have stumbled upon those two, but in the context of having all these digitized, machines can go through and identify similar images that we can find these two images and say, “You know what? The later artist probably copied this earlier one. This is too similar to be coincidence.” So, the power that data offers us to really understand imagery. And this image here, starting in the back left, this is one image selected from each year, so this is 500 years of images. The top left is the year 1500, one image selected. The next year is 1501, 1502, 1503, and on down, so you’re seeing in one image 500 years of how imagery in books has evolved over time. If you zoom into this, it is actually quite fascinating how the styles have changed, to see the introduction of color, to see all these fascinating things like the scientific revolution go through there. Again, there’s really powerful ways of visualizing and understanding society.

But then I want to come to this image. So, what you’re seeing if you look in the lower left, you can see the year, and what this shows is, I applied tools of algorithms to go through all the books published in a given year and pull out every location on Earth mentioned in those books in a given year. What you’re seeing here, and so we’re at 1870, 1880, 1890, watch and in a moment, you’ll notice that the map is getting bigger and bigger and bigger, and then it just pops. It suddenly shrinks. Well, that’s 1923. That’s copyright resuming. And so this map is a very powerful map because it really demonstrates to us how much copyright is impacting digitization and especially data mining, and the conclusions that we can draw from all this data because we know so much between 1800 and 1922, all of this incredible amount of data is available to us, but come 1923 when so much of the digitization efforts have basically ended at that point, we basically loose—we understand more of what happened 200 years ago than we do in the last 70 years. So, this is just something to keep in mind as we think about the world.

But, I also want to bring that same theme to this map here. So, the map you’re seeing at the top there, it’s basically a grid of the world looking over the last three years at any location on Earth where...
there was a large number of “Tweets” sent from during that period. This is an animation by month over the last three years from 2012 to the end of 2014. You’ll notice Twitter expands over time, but then it kind of halts. It kind of freezes in time in the middle of 2013. The bottom two graphs show you basically a scaled 1% sample showing the number of “tweets” per month and the number of unique users sending “tweets” per month, and this is very important because we think often times today in our big data world, we grab for data. We say, “Wow, we’re drowning in data!” Hundreds of, actually in this case, billions and billions and billions of “tweets.” Look at all this data that shows a society, but to never lose track of the fact that yes, we have these incredible volumes of data, but in the case of something like Twitter, this is only reflecting a really tiny portion of the world, and that view of the world has really frozen in time. For folks like me that are data miners, this is really important because social media is not frozen. Social media is just exploding across the world. I think Facebook has now like 1.7 or 1.8 billion users now or something like that, but that data isn’t accessible to us. The majority of Facebook posts are private. The majority of social media as a whole is set to private. So Twitter was kind of the great public experiment. We have all this data to try and peer into society’s soul but to never lose track of the fact that that never really let us look beyond a very small portion of the world, and that view is really frozen. I think this is something that we all too often lose track of it is, who are we really hearing from? What are the views that we’re really seeing across the world?

And this was an example, just looking at this red box there. Some of you may recall last year there was this big battle between Yemen and Saudi Arabia. In Yemen, they fired a missile upon Saudi Arabia. Well, that box there represents Yemen, and all the white dots here are anywhere that any “tweets” were sent in that period. You see that up north of that in Saudi Arabia there are a lot of people “tweeting” there and in Yemen not so much, so we’re really missing half the story. Even as the stories we can tell through data become richer and richer, the comprehensiveness of that is shrinking, and I think that is a critical thing to think about. You think about 20, maybe just 10 years ago, you might be looking at say a gigabyte of data out of a 10-gigabyte data set. Today we’re still looking at a gigabyte, but it is out of hundreds of terabytes of data. The tools that we use, the questions that we ask haven’t really scaled beyond what we were asking a decade ago. It always amazes me that I think it was last year was the 50th anniversary of the “Origins of Dialogue,” and it always strikes me that half a century later the majority of the ways that we interact with information is still through keyword searches. Half a century later, we’re still searching through keywords, and that has always struck me as that is not the natural way to search information and to think about the world.

Even when it comes to things like news, so this is an example. This is March of 2015, and this is looking at—orange are all the locations mentioned in BBC news coverage. Green are all the locations mentioned in New York Times coverage, and for those of you that work closely with the social sciences, you’ll know that there’s a whole area of social sciences, political science, communications, and so on where they hire a large team of humans to look through newspapers and write about what they are seeing there and try to catalog the world. But what we find, of course, is typically that involved looking at something like The New York Times, one single paper, and really data is very powerful because it allows us to really visualize and demonstrate that the world is this incredibly rich place, and no single source gives us a perfect view of society.

And that is why—so what I really want to talk to you today is a project that I call the GDELT Project, which is an open data project supported by Alphabet Jigsaw, formerly Google Ideas, and a number of others. And this project is really about how do we catalog human society? How do we take all of that data that is out there today and try to understand what is happening around the world, specifically in a particular area, so that we can try to bring the world closer together? Because if you think about all the things, if you open up say The New York Times today or a paper here in Charleston, you’re only seeing the smallest microcosm of society. How can we reach across the world and try to understand what is happening in the entire planet so that we can tell stories that aren’t being told, that aren’t being heard?

Like any good project, this begins with data. So, news media, over 100 worldwide news media, print, broadcasts, and web, over 100 languages are processed; 65 of those are live translated. And in partnership with the Internet Archive, we sent them
a list every night of all the online media we find around the world, and they archive them into the way-back machine to preserve online journalism. This is one of the largest initiatives in the world to preserve online non-Western journalism because so much, I mean, take for example the coup in Turkey. Shortly thereafter, almost half of the major media in Turkey disappeared overnight, so being able to really preserve that for perpetuity. Television—there’s another collaboration. We’re looking at that. Academic literature—this I think also really demonstrates the ability of data miners to collaborate with publishers. So, this is a project—among others, JSTOR provided access to a lot of its holdings to be able to data mine and to be able to look through academic literature, to be able to understand what have social scientists and humanists and ethnographers and linguists and so on, what have they really written about the world over the last half century? What can we understand about that? How was the output of academia over the last half century? What can we understand on, what have they really written about the world humanists and ethnographers and linguists and so on, what have they really written about the world over the last half century? What can we understand about that? How was the output of academia compared to say the output of the news media? You know, how are these really combining? Books, of course, and then imagery, which I will come back to in a second.

Of course, if you look across the world, most of the world’s media and most of the world’s information is in a language other than English. What you’re seeing here is a “dot” at every location which GDELT monitored information from or about over about a six-month period of last year. It is color-coded by the primary language. Gray is English, and you’ll notice there’s not a lot of gray on this map. So if you’re only looking at English-language material, which is what most data mining does and what most American scholars tend to focus on, you’re missing most of the world here. Yes, there are plenty of news sources and other materials published in France that are published in English, but those don’t reflect local views or local events or local things that are happening there. So, just always remember how important it is to look at information. When you’re trying to understand the world, you have to turn to all the world’s languages, and in this case using machine translation. Now machine translation is far from perfect, but it’s good enough. For example, I don’t speak a word of Thai. So, if I see a Thai language article, I can’t tell you if this is say a local cricket match score or if this is an anti-government protest. Being able to leverage the technologies that we have today to really try to understand the world, you can imagine bringing in all of this information from across the planet, and anything that is not in English translating that in English, what could you do with all that? Why do you want to do that? What is useful about all this? Well, two things.

The GDELT Project looks at two things: It takes—one, it tries to understand the physical events. So, what is this newspaper article telling me? Is this reporting on a riot, a protest, a coup, a peace appeal, a diplomatic exchange? So, that is the event data set. So much of what we are interested in and around the world is not physical events. It is narratives. It is emotions. It is beliefs. Those that work in the humanities world know how important that is there. It is not enough to say, “Hey, a coup took place in Turkey.” We all know that. What we care about is how are people reacting to that around the world? What are the emotions and the narrative beliefs?

What can we do with this? Well, for a physical event data set, if you take a large fraction of the world’s news media and you process that, and you make a list of every protest reported around the world over the last day 40 years from the data set, you can take countries and actually plot the intensity of protest activity and ask questions, like when Greece was having its issues, what was that, a year ago? I saw this article, I think it was on CNN, that said, “Greece Undergoes Worst Violence in its Modern History” and to being able to look back and say, “Yeah, there was certainly some unrest but nothing compared to five years ago, and certainly nothing compared to times before that.” So, really being able to contextualize things because, again, we have such short memories and being able to really understand the world around us or being able to do automated alerts. To be able to say not many people were watching Burundi in December of last year, but to be able to say hey, all of a sudden something was happening there that we need to pay attention to. So, being able to put that in front of say peace builders to be able to—again people on the ground, obviously people in Burundi know what’s happening there, but being able to tell that story to the world. Because when something happens in Turkey, the world listens. If something happens elsewhere in the world, it’s not making the front page of say The New York Times, so how do we tell these stories? How do we tell what’s happening?

And I won’t go into this slide, but you can do things like actually look at the cycles of world history. We can actually build mathematical models over
hundreds of millions of events that have taken place across the world and actually find similarities there; cycles, patterns that repeat themselves that allow us to start really peering into the soul of society. We can do things like map. Now this is very old data. That’s why it is very sparse. This is I think from 3 years ago, but being able to actually map the world, actually watch the world go by moment by moment, to say give me a map of all the protests happening right here in the world. What are the grievances? What are the things that are driving society? Is the world becoming more democratic or more autocratic around the world? Being able to ask these questions, to say you know what? We’ve had things like news media and other data sets. These are not new things. We’ve had them for very long time, but for the first time, we had the data to be able to ask questions of that. To be able to map things beyond, things like emotions, but to go beyond positive and negative. This is actually a map of anxiety, and this is the big trend mine. The surge you’re seeing there is the U.S. government shut down. This becomes very powerful because this was the case where there wasn’t a lot of positive or negativity because you had a lot of people say, “Hey, great! I get a month vacation! Fantastic!” You had other people say, “This is horrible! Our government is shutting down! This is awful!” So, in terms of positive/negative, both sides, there weren’t—it was sort of like the election today, who is going to win the election? Well, depending on which candidate you’re going for, it’s a good or bad thing. But anxiety is very high because either side you’re worried about things so being able to really ask deeper questions about society and being able to look at things like Ebola coverage.

So, this is a map, a timeline of coverage of Ebola in American television news, and you can see that thousands of people are dying across the continent, and there is no news coverage. Television is never mentioning it. It is not until the two Americans get it that suddenly it is worth American media covering it. Television is never going to cover it unless there’s some huge ivory seizure. I don’t really see much in The New York Times each day. Well, that’s true, but local media in local communities do cover this every day, so always remember that even if we don’t see things in the information that we consume, then just remember that globally there is so much information that is out there, and the dark map, the network diagram in the center, was something that BBVA, the big bank in Spain, they wanted to see how was Russian media portraying the sanctions, the economic sanctions against Russia. So being able to say, “Well, how is the media within a country portraying something compared to what we’re seeing out here?”

And in particular, I want to show this map. This was something. I wanted to map global happiness. How happy or sad are people of the world? This is a question people have asked for eternity, but I said, well, what if we take, in this case 200 million articles published last year in over 100 different languages, 1.3 billion mentions of locations on Earth, and about three quarters of a trillion emotional assessments off
of that. What I did was say, well, every article worldwide that mentions Paris, what is the average happiness or sadness of all of those mentioned across the entire planet and do that city by city across the world? You end up with a fascinating map like this. There are so many fascinating stories here. You also start reflecting media. Certain country’s media reflects more of a negative tone toward things, but in particular in Europe, it’s a little bit hard to see on the screen, but you see these really strong negative tendrils go through Europe. That’s a domestic reaction to the refugee crisis that is occurring there. So, being able to see how domestic countries are reacting to world events and just seeing that on a global stage is very powerful. By the way, I should mention that was one line of codes, so that was a tool that Google makes called Big Query. It’s a super database platform. One single line of SQL query, 60 seconds. One minute later and you’ve processed that much data. So, really we’ve moved to the ability where if we have a question, we say I wonder or what if? The fact that we can actually answer those questions or be able to peer into that media and say who are all the people talking about a particular issue? Like, CCS is a clean energy technology and being able to look inside of that and say what is this environment look like? What does this landscape look like? And this is something I want to conclude on.

So, this is a project—Google, like many companies, is doing deep learning, neural networks, artificial intelligence. They are doing some amazing work on having AI systems. They can look at images and catalog what are the objects in that image? What are the activities in that image? Does this image depict violence? What are the logos, the locations? Can we look at the background of that image? And look across the planet and say can we estimate the location of this image? And that is very powerful because we can use this to do incredible things. What you see in the upper left, I asked it for all the images of trash around the world. This is actually something I will be putting out shortly. This looks across the world in real time and tries to estimate basically environmental conditions, so litter conditions, air pollution, and so on across the planet but in particular looking at things like flooding. One of the things that we do is when a natural disaster like Hurricane Matthew passes through, being able to take all the images emerging from those areas and say, we don’t want just presidents at podiums. We just want the images that show the destruction so we can send those off to first responders so they know what’s happening across these areas. So again, being able to move beyond text to really understanding the imagery of the world. And the fact that we have the tools today to look through, and so far, we’ve processed 175 million images since this past January from every country on Earth in real time. The ability to use tools to really try to understand the narratives of the world, and I want to conclude on this image. This image really, by the way, this is a real image; this is not Photoshopped. This is the NOAA science ionosphere. It is really cool. It is a 6-foot acrylic sphere hung from the ceiling with projectors around it. This is one of my data sets on it. This really kind of summarizes what I think is so powerful about today, that we have so much data. We have the tools—we’ve always had more data than we can deal with, but we have the tools for the first time, the machines, the algorithms, etc., to make sense of that data, to let us peer into the soul of society and really understand it in ways that we’ve never been able to do before. Thank you so much.
The Long Arm of the Law

Ann Okerson, Senior Advisor, CRL
Mark Seeley, Senior Vice President and General Counsel, Elsevier
Bill Hannay, Partner, Schiff-Hardin LLP

The following is a transcription of a live presentation at the 2016 Charleston Conference.

Ann Okerson: Good morning, and welcome to the seventh, I can’t believe this is the seventh annual session, of the Charleston Conference Long Arm of the Law. And we want to especially thank once again our guest star, Mr. Kenny Rogers (playing Kenny Rogers song). Next time we’ll have everyone sing it, right?

So, with those words of caution, I wanted to say that two or three weeks ago I sent out a note on LIBLICENSE-L asking people to send me what they thought had been some of the key developments in legal issues related to libraries or publishing over the course of the year. I got a shorter list than I expected, but I’ll read it to you. The Section 108 meetings happening in Washington, DC, the copyright office at the Library of Congress, and then, of course, the recent shakeup, Sci-Hub and article sharing, the Georgia State case which seems to have more lives than a cat, the American Disabilities Act and websites, and the right to be forgotten. Those are some of the items that came in. As you know, the legal issues that relate to us are profuse. They are numerous and ever-changing and never-ending.

So, what we thought we would do this time would be a little bit different. We invited Mark Seeley, who is Lead Counsel at Elsevier, to be our first speaker here, and I asked him to talk about the day in the life of a legal attorney in a publishing company, a large publishing company, obviously, and he is going to do that, and then we move on from there. Mark is Senior Vice President and General Counsel of Elsevier. He splits his time between Cambridge, Massachusetts, and the Amsterdam headquarters. He leads an international team of publishing and sales lawyers, and the Global Rights and Permissions Team also reports to him. He is also on the Board of Directors of the Copyright Clearance Center. He chairs the Copyright and Legal Affairs Committee of the International Association of STM Publishers (Science, Technology, and Medicine), and he’s a member of the AAP Copyright Committee. He regularly contributes to papers on copyright issues and best practices. He is a frequent speaker on copyright. His education is Thomas Jefferson College, Grand Valley State University of Michigan, for the BPh in Literature; Suffolk University Law school in Boston for the JD, and he’s a member of the Massachusetts and New York Bars.

Our second speaker Bill Hannay is known to us all. Bill is a lover of libraries; although he’s an anti-trust attorney, he’s long loved libraries. I met him back in the early ’90s when I worked with ARL. He likes to come to Charleston, and he always composes a song for us, so I think he will not disappoint. Bill regularly represents corporations and individuals in civil and criminal matters involving federal and state anti-trust law and other trade regulations. He’s an Adjunct Professor, teaching courses at IIT, Chicago Kent Law School in anti-trust, intellectual property, and is the author or editor of several books on anti-trust and IP property law, including Corporate Counsel’s Guide to Unfair Competition, soon to be published by Thomson Reuters West Publishing. He is a frequent lecturer at the Charleston Conference. He’s active in the American Bar Association. He’s currently co-chair for the Joint Editorial Committee for International Law. He served as Assistant District Attorney in the New York DA’s office, and was a law clerk for Justice Tom Clark on the U.S. Supreme Court. He’s a graduate of Yale College and Georgetown University Law Center.

Now, the format that we’re going to follow today will be a little bit different. Mark will speak first, and we will have then a few minutes after that for comments and questions for Mark, because his presentation is of a different sort than Bill’s. And then after Bill’s, we will have another chance for more comments and questions to Bill and to both of them.

Let me introduce Mark, and thank you very much for coming.

Mark Seeley: Thanks, and good morning everyone. Actually, it is not my first time in Charleston, but it is my first time at the conference, so it’s great to be
I think everyone in this room will know Elsevier as the publisher of many journals, and that is certainly an important part of our business, but even on the publishing side, we also publish a fair number of books and databases and the like, and then increasingly our businesses is focusing on questions about analytics and services. Some of that is based on the scientific or research-intensive side of our business, so building on the content that we are developing in terms of things like databases and how can we turn those into analytic services to help institutions look at their output of research activity. But also, we do a fair amount on the health side working with hospitals and healthcare providers and insurers to look at the effectiveness of their activities. There is a lot of writing which is not exactly scientific writing; it is more about practice and medical practice. We also train and test a large number of students. We do something like 750,000 tests are done online every year by the Elsevier business through the old Mosby business or HESI Business, as it is now known. So, you see there is traditional publishing, but there is a fair amount of new analytics.

And the implications of that are that we have to do a lot of different kinds of activities, some of which are quite traditional in terms of things like the publishing contracts of one kind or the other, but we are also increasingly doing distributor and agent agreements, technology and procurement contracts of one kind or another. So, there is a large scale of contracts here, but there’s also a large scale of expertise that we are asked to provide, things about procurement problems, compliance issues like data protection and privacy or antibribery, and the like. So, my challenge is how to do that with the department that we have.

Now because of the RELX corporate structure, we do have a central corporate legal team at the RELX level, and they provide some centers of expertise for us in terms of mergers and acquisitions, some work in the patent area, labor and employment issues, which is incredibly important when you have 7,000 staff around the world, and things like compliance and data protection and privacy. The Elsevier legal team that I manage directly is 19 lawyers, and the numbers are reasonably split between the U.S., Europe, and APAC; that is, there is almost as many lawyers in Europe, including the U.K. for the moment, as there are in the U.S., and the U.S. has our litigation team. Otherwise, if we were just looking at the business supporting lawyers, the numbers would be much more equal, and then the
numbers are increasing in Asia as the business is developing there. Then we also have a sizable team of rights and permission folks, paralegals, administrators that are part of this.

The way that we have organized ourselves is we have four teams within the legal function. Two of those are very focused on traditional publishing, one on journals and one on books and databases. One is focused on sales issues. One is focused on technology and procurement, and we have our litigation team as well, so you could say that we have five teams altogether. We have regional General Counsel, so there’s one regional GC that really supports the European and APAC business, and one in North America, and they’ve got a variety of responsibilities, including liaison with management teams in their general areas.

Generally, the way I think that any company has to look at the balance between business needs and, by the way, this is true at any institution or university also I believe. It’s not really unique to businesses. How do you balance those needs with the resources that you have, and we think of this very much as all about triage. It’s about managing the resources in an efficient and effective way and trying to think about those large numbers of contracts, for example. Are there ways to do more of that online? Is there more automation that is possible to be done, or generally speaking, can we provide more tools, more self-help procedures, if you will, for folks in the publishing and the sales side of the business? That is really how we look at the business and how we try to help the business both do things efficiently, but also through that, efficiency gives us a bit of resource to help on the more strategic dimensions. So, on complex negotiations, on sort of looking ahead and down the line in terms of some of those questions about technology and analytics, being efficient means that you can do more of that work as well, which I think provides a better resource for the business and the company. That is a little bit about the business and a little bit about the legal department. Enough about that. Let’s talk about me.

These are, I have a number of key objectives, don’t we all, every year—I’m going to talk about three of mine for the year 2016, and I think some of them won’t be surprising at all, so copyright and public policy as I said is my bread-and-butter, something I deal with on a regular basis. Compliance, of course, we have to comply with laws and regulations. Sadly, laws and regulations are increasing rapidly all the time, and we cannot rely on the idea that the regulations are only relevant if they are in the US or the UK or the Netherlands. There are lots of regulations that are happening around the world, and some are just as difficult and just as intense as they may be in the States and Europe. And I’m also going to talk a bit about the collaboration and analytics question.

In the area of compliance, it is always about risk assessment, of course, and providing advice and support in terms of investigations. It is always a lot about training and identifying what are the rules of practice that we’re going to follow. It is easy to say that you have to comply with law around the world. It’s harder, I think, to provide a real set of procedures and controls that we think generally ensures that the business is operating lawfully. One example that I give here is the antibribery program. A few years ago, the UK joined the US in having quite stringent regulations on antibribery, and the difference between the US and the UK was that the US tends to focus on government agencies of one kind or another. UK law was much broader, and the U.K. law really required that you really know the people that you’re dealing with in terms of distributors and agents and the like, and particularly that you have some responsibility for those entities out there that are actually acting on your behalf, so, from a legal perspective, acting as an agent. What we had to do several years ago was we had to stand up a program by which we assessed the more than 100 distributors and agents. And by the way, when we started the program, we discovered that we had something like 400 distributors or agents, so part of the program was surely we don’t need to have 300 or 400 agents and distributors around the world. Let’s focus on the key ones and really drill down. So, we did a lot of this initial assessment ourselves. What we decided, and this was a lot of work for us to do, the due diligence process, of course, involves a little bit of questionnaires of the agents but then doing some searching, using some LexisNexis tools and others, to see whether, in fact, the person or the party that you’re dealing with seems to be ethical, seems to be dealing with their customers in an ethical way where there are no reports that that agent is somehow involved in bribery or other unethical issues. So, the requirements under the U.K. law is that you do that kind of an assessment on a periodic basis. Last year, as we were coming up on the renewal that we set ourselves to relook at all agents and distributors around the world, we decided that it would be clever
to do that efficiently, and we set up an industry, an independent industry bureau, to conduct due diligence, and we shared that cost across several publishing houses. The simple idea there is that if 30 publishing houses are asking one agent to fill out 30 questionnaires and going through a due diligence process, why not do it once and do it more effectively? So, that is an example at the kinds of issues we look at on the compliance side.

I think the copyright issues you won’t be surprised that there is a strong focus for me, and frankly for the entire legal department, on these issues, and I think it takes, it is really about three dimensions. One is our internal policies. We’ve always tried to look very carefully and not to make sort of automatic leaps of judgment about what is right and what might not be right, and the issues there are particularly acute on the journal side of the business, where after all authors are looking for visibility and public claiming of their inventions and discoveries. Somehow the publishing world involved in journals has to find a way to live with that desire for visibility, while at the same time, particularly on the subscription side, preserving a business model. So, how do you balance these things? There’s a lot of internal policy discussions that we have, but we continue to sort of manage the copyright issues, rights and permissions, clearing permissions and the like. Even in an OA world, there is a fair amount of issues about copyright, for example, Creative Commons licenses and all the flavors of those licenses and which is more appropriate.

We are also looking on the enforcement side, so we look at sites that are using content, and we try to identify the best way to reach out to them. We have been focusing a lot through the STM Association on a set of sharing principles. Again, it’s the idea of what can we support in terms of visibility balanced with a need to maintain a subscription business, and then we are also looking to see what type of issues are going on legislatively, so what are the copyright revision efforts which are being looked at around the world? I think at the moment this is most acute in Brussels. The European Commission just released a few weeks ago a new document called “The Digital Single Market.” Of course, it is very much at a proposal stage and will go through lots of changes legislatively and elsewhere, and the key issue that we’ve been looking at here is the question of text and data mining rights as an exception to copyright. And our key point here is to try to preserve the commercial market, which is pretty viable and which is growing nicely. Think about the pharmaceutical industry, for example. They’re very interested in the question of text and data mining, not only of published content but also of their own content as well, and they’re looking for tools and services that help them do that.

The third thing that I thought I would talk about is this question about collaboration and technology. As we look at our business, what we think is that the future, of course, has a strong technology focus, and, to some extent, the future is about what kind of answers can you provide ultimately, so it’s not just about doing research. It is also about finding ways to work with technology and big data, and I know there’s been lots of discussions about big data over the past couple of days. How do you provide the kind of combination and collaboration of technology and content to provide better answers and better information for researchers and for medical practitioners? The questions that we have as we reach out to third parties to think about doing these kinds of collaboration projects, some of them are not unusual. In almost any partnership or collaboration, you’re always going to have differences of view between the respective partners as to the value that they are bringing to the party, so it won’t be a surprise that from a content perspective we think the content that we have worked on, both on the science side and the health side, is pretty valuable and pretty useful, and we think of content as sort of being king. Surprisingly, the technology vendors have a completely different view, and their view is all about delivery. It’s all about solutions, and content is kind of a commodity, so it’s easy to see how there can be collisions of interests and disagreements.

Part of the exercise here is to try and figure out what kind of approach works, what different types of technology collaborators, and also to be thoughtful about questions about what types of intellectual property rights, and actually, I don’t think it is about IPR. I think it is about intellectual content that is being thrown off as a result of these kinds of projects because it is not the sort of traditional IP rights that we are used to. It’s probably not copyright, and it is probably not patent, and it is definitely not trademarks, and it’s not trade secrets, if you’re going to talk about it a lot, so what is it and how do you protect it?
Those are some of the key personal responsibilities for this year that I am responsible for, and then here is the kind of the final part of the discussion and the last slide here which is “The Day in the Life” issue, and for me, this is actually an excerpt from the week of the 3rd of October. Now that is important because in the publishing world that means that is about two weeks away from the Frankfurt Book Fair. As probably many of you know, the Frankfurt Book Fair is one of the largest international gatherings of publishers and distributors and agents and even some librarians that happens around the world, although London Book Fair would complain about that characterization. So, the thing there is that because all the publishing houses and all the trade associations, well many of them, are meeting that week in Frankfurt, of course there’s a lot of preparation for those meetings. So, what are the key issues that are being discussed in terms of copyright issues, copyright revisions and the whole question about text and data mining rights and the digital single market was critical there. So, a lot of preparation and a lot of discussions within the trade associations and in individual one-on-one discussions with publishing houses. Some related discussions about technology because I think as actually have been discussed quite a bit here in sessions at Charleston this year, there’s a lot of issues about both better accessibility and better security, and are those two things completely in conflict, or are there ways to improve accessibility and ease of use while ensuring the security is also there? And publishing houses are looking at those issues as well.

On the collaboration side, we did an in-person workshop. I gathered together some senior managers at Elsevier with an external lawyer because, of course, we won’t have expertise in all of these areas and all these issues and will always rely on outside counsel to provide some particular expertise on particular points, for example anti-trust issues. We definitely would talk with outside counsel about those issues. So, we did an in-person workshop and tried to work through some of those questions about valuation and asked that in these combinations. I had a couple of compliance issues, so it surprised me to learn that if you operate an online job board from the UK that you are considered to be an employment agency. I had no idea that this was the case, and it struck me as completely wrong, and what I gathered is that most online job boards that operate in the UK do not regard themselves as employment agencies, so they kind of tend to disregard it. But nonetheless, that was what the law said, so we had kind of work through what the implications of that were and how we could actually operate the system going forward.

We also had some clients’ investigations into APAC countries. One was a result of an internal whistleblower, and one was a result of a government agency, and here all you can say is that all the training in the world and all the best practices and a code of conduct, at the end of the day, people may be motivated by things other than the best business ethics, and you do need investigations, and you need, frankly, penalties to really ensure that a compliance program really operates and works well. I had some administrative things going on as well, so we were looking at the Books Contract Automation Project, which we’ve been working on this year and which will be standing up next year in 2017, but we also had some corporate organization questions. We had some changes in directors. When needed to look at the slate of directors for the Dutch, UK, and US entities. These are not the publicly traded entities but the operating entities, and we had a discussion with our tax team about some assets that are owned by a European entity which is no longer terribly active. And then finally I actually did some publishing things, so I sat—I’m one of three members of the retractions panel inside the company that looks at the journals and books in retraction and removal proposals, and this often gets us involved in discussions with the external journal editors about what they’re proposing, how they’re proposing to do it, and make sure that we are well on. So, that was a day in the life. Thank you.

Bill Hannay submitted a written paper for his portion of the presentation, included below.

**An Update on the “Right to Be Forgotten”**

As you may recall from prior “Long Arm of the Law” presentations, the European Union vigorously protects privacy rights. Twenty years ago, the European Parliament and the Council of Europe adopted the “EU Data Protection Directive,” i.e., Directive 95/46/EC of 24 October 1995. It protects individuals with regard to the processing of “personal data” and the movement of such data.

What is personal data, you may ask? It is any information relating to an individual, whether it relates to his or her private, professional or public
life. It can be anything from a name, a photo, an e-mail address, bank details, to posts on social networking websites, medical information, or a computer IP address.

Two years ago, the European Court of Justice had down a landmark ruling in May 2014 that EU privacy law required Google to take down (or “de-index”) negative information about an individual citizen of Spain, Sr. Mario Costeja. See Google v. Agencia Española de Protección de Datos, Case C-131/12.

On May 13, 2014, the ECJ held that Google (as an operator of a search engine) is obliged to remove from the list of search results any web pages links relating to an individual if such information is “irrelevant” in relation to the purposes for which the data was collected or processed and in the light of the time that has elapsed.

In short, the ECJ required a “balancing” of the legitimate interest in access to information and the data subject’s fundamental rights.

The court’s decision opened a floodgate of privacy requests from other EU residents. In that past two years, Google has received a half million requests to remove information and has complied with 43.2% of them. While many applaud this development, there has been some fear among historians and librarians that the role of libraries in preserving historical records is being impaired.

The 1995 EU Data Protection Directive will be replaced in 2018 by the General Data Protection Regulation, but the new rule will not cut back on the “right to be forgotten.” EU citizens will still be able to request data custodians like Google to remove negative information about individuals, but there remain limits on it, as the European Parliament stated in approving the new regulation in 2014 and the Council of Ministers repeated in endorsing the regulation in 2015 FN/:

The right to be forgotten is . . . not an absolute right. There are cases where there is a legitimate reason to keep data in a database. The archives of a newspaper are a good example. It is clear that the right to be forgotten cannot amount to a right to rewrite or erase history. Neither must the right to be forgotten take precedence over freedom of expression or freedom of the media.

The latest controversy about the right to be forgotten is the ruling of the French data protection agency (CNIL) in September 21, 2015, now on appeal to the French courts. There, the CNIL ruled that Google must take down or “delist” results on all of its extensions, including its U.S. portal, Google.com. The ruling is not just limited to Google’s European ones (e.g., .fr; .es; .co.uk). Thus, the French ruling would directly affect searches done in the US.

The International Federation of Library Associations and Institutions (IFLA) is a strong voice urging restraint in applying this privacy right. Most recently, in an October 2016 letter, IFLA urged the French courts to reverse the state agency and not to expand the right beyond national borders.

Can the ADA Spell the End of MOOCs?

On August 30, 2016, the U.S. Department of Justice formally notified the University of California at Berkeley that it had violated Title II of the Americans with Disabilities Act (ADA) by making free audio and video content available to the public on YouTube and iTunes and in MOOCs . . . but not making that content accessible to the deaf and blind. The DOJ advised Berkeley that it must modify its free offerings and “pay compensatory damages to aggrieved individuals.”

In September, Berkeley issued a statement that it is—in effect—between a governmental rock and a fiscal hard place, unable to afford the cost of restructuring the programs. It may, therefore, have to remove the content from the public. Sadly, this is a no-win situation.

And Berkeley is not alone among schools that have been sued by the DOJ for ADA accessibility violations: 25 others have too.

Where will it all end? It is hard to say at this point. Perhaps the Trump Administration will take a different view of the situation.

Georgia State—E-Reserve Case

As you may recall, Georgia State University became the target of a copyright suit for allowing professors to designate portions of books and periodicals to be copied by the library, scanned, and put on “electronic reserve” or compiled into “electronic course packets.” Three publishers (Cambridge University, Oxford University, and Sage Publications)
sued, alleging that substantial portions of 6,700 works had illegally been copied and transmitted to students for some 600 courses at the school.

After discovery, the cast proceeded to trial, and in 2012, the district court largely ruled for Georgia State, holding that it was “fair use” for the university to electronically copy up to 10% of a book or even a whole chapter. \textit{Georgia State University v. Becker}, 863 F. Supp. 2d 1190 (N.D. Ga. 2012) [Evans, J].

In 2014, the U.S. Court of Appeals in Atlanta reversed and ordered the trial judge to take another look, using a more nuanced analysis. \textit{Cambridge Univ. Press v. Patton}, 769 F.2d 1232 (11th Cir. 2014). Significantly, the appeals court held that the nonprofit, educational nature of the university’s use of the material favored a “fair use” finding.

Publishers were horrified. They looked at this sort of wholesale copying as undercutting the entire ecosystem of academic publishing. They hoped for a better result on remand, but that did not work out for them. In March of 2016, the trial court again ruled in favor of Georgia State after taking a second look. The court largely tracked the same logic as before.

Where will it all end? Spurred by the apparent success of Georgia State, other colleges and universities have adopted similar e-reserve and/or e-coursepacket approaches. Publishers have fought back, filing similar cases against U.S. universities, including UCLA, and against foreign institutions, including York University, Delhi University, and in New Zealand. The jury is still out, but the publishers have so far not done well in the Indian case.

\textbf{Delhi University Photocopying Case}

In September, a trial court in India ruled against publishers in an even more blatant case of copying, one where the university worked directly with a photocopy service to make hardcopy course packets for sale to students. See \textit{University of Oxford et al. v. Rameshwari Photocopy Services et al.}, CS(OS) No. 2439/2012, High Court of Delhi, Decision dated 16 September 2016. The trial judge stated:

That, in my view, by no stretch of imagination, can make the [photocopy shop] a competitor of the [publishers]. Imparting of education by the defendant . . . University is heavily subsidized with the students still being charged tuition fee only of Rs. 400 to 1,200/- per month. The students can never be expected to buy all the books, different portions whereof are prescribed as suggested reading and can never be said to be the potential customers of the plaintiffs. If the facility of photocopying were to be not available, they would instead of sitting in the comforts of their respective homes and reading from the photocopies would be spending long hours in the library and making notes thereof. When modern technology is available for comfort, it would be unfair to say that the students should not avail thereof and continue to study as in ancient era. No law can be interpreted so as to result in any regression of the evolvement of the human being for the better. (Pg. 84)

Social advocates hailed the verdict, saying the court had correctly upheld the supremacy of social good over private property. Students had rallied behind the photocopier, saying most of the books were too expensive.

The publishers plan to appeal, arguing that the trial court’s approach goes far beyond any reasonable interpretation of the exception in the copyright act for educational copying.

Stay tuned for next year’s updates of these fast-changing legal areas.

\textbf{References}


69 Charleston Conference Proceedings 2016
Update on Industry Trends and Issues

Gary Price, Co-Editor, InfoDOCKET
Rick Anderson, Associate Dean for Collections and Scholarly Communication, University of Utah
Maria Bonn, Senior Lecturer, University of Illinois

The following is a transcription of a live presentation at the 2016 Charleston Conference.

Gary Price: Good morning, everyone. My name is Gary Price. I’m here with Rick Anderson and Maria Bonn. We’re going to try to spend the next 45 minutes or so sharing with you. Not only are we going to share some open web resources, and we’re going to discuss them to some degree from a policy standpoint. All of the resources that we are going to talk about and much more are on this webpage: bit.ly/abpINchs, and I’ll bring the URL back up at the end of our presentation.

So, first I wanted to focus on some open access and look at it from a copyright intellectual property standpoint. The two resources in this area that I wanted to share were one: It’s called BASE. Have any of you heard of this? A couple people, but not many. This comes from the Bielefeld University in Germany, and as of about a week/10 days ago, this database now has information on over 101 million articles and other documents from over 4,779 content sources. In addition to being able to search all this data and offer links for the full text, they also, if you like to create your own search tool or pick and choose, here is a database of all 4,779 targets. If the metadata is available, there is the advanced search page where you can narrow by terms of use, terms of reuse and licensing, also access, limits to different type parts of the article, again if the article metadata is available. So, this is BASE from Bielefeld University in Germany.

And the past week, a new resource based on this search tool and other resources became available, and this is called oaDOI: Leap over tall pay walls in a single bound. So, this is leveraging all 101 million articles in that database and often preprints. You might’ve noticed very quickly that when you go back to search the main home page you can give open access resources a boost in the relevancy ranking algorithm, but this makes it even simpler for those people who have oaDOI. You can basically just plug the DOI in, and it will tell you if there is a preprint open access version of the article available using that BASE search tool and other resources. It’s not perfect, but, I think as Rick said in our little meeting yesterday, this is kind of a free version of Sci-Hub, taking some of those same concepts. So, that makes BASE: API available and oaDOI: API available. So, I was going to ask my colleagues to comment a little bit on these two. Rick? Maria?

Rick Anderson: I actually think my reaction when I saw this that it’s an ethical version of Sci-Hub.

Gary Price: Yes.

Rick Anderson: But, the thing that occurs to me when I look at tools like this is there’s a lot of conversation among us in libraries right now about how and whether we should be integrating open access resources into our workflows. Shouldn’t we be cataloging these things that are available and making them easily findable in our workflows? Shouldn’t we be cataloging these things that are available and making them easily findable in our catalogs or other web-based tools? And I look at things like this, and I think, “No, we should let somebody else do it.” We should have a link to oaDOI or BASE or whatever. Why on Earth would we all—and this really goes back to the question we’ve been asking about cataloging for 20 years—why on Earth should 500 libraries catalog the same book, when one of us or two of us could catalog that book and share the record with others? So, what I find kind of exciting about these tools is that to me they indicate the possibility of a way forward that makes this stuff radically more findable without me having to devote staff time to it.

Maria Bonn: I would say the danger of doing something like this with Gary because he sends you a list of links and says, “Let’s talk about these” and then you stay up all night. So, this is leveraging all 101 million articles in that database and often preprints. You might’ve noticed very quickly that when you go back to search the main home page you can give open access documents a boost in the relevancy ranking algorithm, but this makes it even simpler for those people who have oaDOI. You can basically just plug the DOI in, and it will tell you if there is a preprint open access version of the article available using

Copyright of this contribution remains in the name of the author(s).
https://doi.org/10.5703/1288284316503
these things? But I’m also an educator. I teach at the School of Information Science at the University of Illinois, and when I look at these, I think, well, what should I tell “the kids”? Some of the kids are my age. What should I tell “the kids” about this? And what should they be telling library users about it? How do they lead library users to this kind of tool? And I agree with Rick that when I was first looking at these I wanted to get on a line with our library director, who I happen to be happily married to, and say get this into the catalog or get it onto the library web page, but it needs to be a kind of seamless integration so that it shows that these are resources in which we have an investment, even though it’s not financial investment, but that can be discovered along with the things which we pay for as well, and these tools are there to use. How do we integrate them into our systems and services is a good point.

Okay, quickly, the thing that struck me, and maybe this comes into the arena of what should the kids be telling library users, is how do the people who host this kind of thing find it and learn how to use it? I think a lot of our principled advocacies on open access comes from a sense of need to help all of the information poor, the people who aren’t students of higher education, and if they are they can’t afford access to all those resources due to a pay wall. So, this oaDOI, I will show this to my students; they will find it useful, although they will find that it is kind of a hassle with the proxy server sometimes, but I can go straight to it this way. But, the proxy server is there for them. But, the kind of person that we like to hold up as benefiting from open access would be the mother below the poverty line who is looking for information on her son’s sickness. Is she going to know about oaDOI—I’m sorry, I’m so excited about this tool. It will be great to use. But, it may be one of things that we need to be thinking about it as our responsibility as librarians to promote and share these tools in the appropriate ways at the appropriate time. That’s it.

Gary Price: Let’s talk a little bit about multimedia content on the open web and tools to access and make better use of it. One is one of my all-time favorite databases that unfortunately C-SPAN doesn’t have the resources to make available and promote like they should. This is actually based, the C-SPAN Video Library is based at Purdue University, and as of this moment, there are over 226,000 hours of video in here. Basically everything that has ever aired on C-SPAN, all of the author interviews, all of the rallies, all of that is available in here. In addition to that, in most cases, I’ll just use Mr. Google himself, and the presentation I’m searching for a person, I could look all the appearances that Eric Schmidt has made on Google all on one page. So, in addition to that, I can either find the video like I just did using metadata, or I can search every single word spoken in most of those 226,000 hours. They allow you to keyword search the transcripts of just about all that material and find it, or you can find the video and then search for words in the transcript this way. In addition, for sharing with others, a professor or student can, for example, create a customized clip. Move the levers back and forth, say from whatever moment to whatever moment, get a direct URL to that specific clip hosted on the C-SPAN server, share it with others, and they are done. So that’s 226,000 hours, thousands of hours of author interviews, political rallies, and, of course, congressional hearings and the like.

Another resource I want to share that’s relatively new is called audiosear.ch. This is a podcast search engine, and as of now, they’ve got 24,000 episodes of podcast. This is both a discovery tool for podcasts, and in some cases, you can also do transcript searching. They’re using technology developed by a nonprofit. In the last year, they’ve just made a deal with the Digital Public Library to work with some of those organizations, but popup archive works with organizations to take audio and create transcripts, speech to text. So, this is also a company that if you have a lot of audio, and you want to create transcripts to make it more easily discoverable and searchable, this is an organization you might want to take a look at. So, that’s that.

So, those are two examples of multimedia search. It’s coming on fast and strong, and now with a relatively low-price organization like pop-up archive out there to allow organizations to create searchable transcripts, I think we’ll see even more of it.

Rick Anderson: So, as a very impatient person myself, whenever I get referred to a video, a little part of me dies because I don’t have 25 minutes to wait for this guy to get to the point, right? And so, for me, a tool like this that C-SPAN is offering is really a game changer, the ability not only search but to then pull up a transcript of what was said in the video. For the work that I do, this is just huge, and what I hope when I see this is that many other purveyors of audio and visual content will either make use of this backend, or if they
want to come up with their own version of it, but being able to access C-SPAN videos is not really a game changer. You could pretty much find most, if not all of these, on YouTube, I think, but being able to search the transcripts is huge.

**Maria Bonn:** In our profession, as you all know, we talk a lot about information literacy, and that generally involves a component of evaluating resources for their reliability, their credibility, “Does this come from a good source? Can I trust this?” Something we worry a lot about, and as I look at the kind of wealth of resources that become available through search tools like this, I mean, they’ve been out there, but now they’re easier to get at so you can get more of them, right? How many thousands of clips, right? I start to think about how do we then further our thinking about information credibility to be able to encourage users to evaluate the credibility of multimedia, and it seems like we’re pretty comfortable with text kind, but that multimedia may add another dimension to our assessment about credibility and reliability. What sort of skills do we need to be equipped with as librarians who are pointing people to resources and what do we want our users to be equipped with? So, I think about those complications when I don’t get lost searching for my dissertation topic that I wrote 30 years ago. If I’d had this then, it would be really cool. And I also think in this goes back to the previous tools, think about terms of use and copyright considerations—do we assume that the service providers have taken care of that, and if we can find it and see it that we can use these and be in ethical and legal compliance? How do we find out when we pull up one of these resources? What’s the trail that leads to the information that we need to make that determination? So, I think both way cool and more things to think about. More problems.

**Gary Price:** Web archiving continues to be a very popular topic and something people are talking about. In the past couple of weeks, this tool funded by Mellon went into version 1.0, webrecorder.io, create high-fidelity interactive recordings of any website you browse. This is a free service. So, I won’t take the time to login, but basically, I’ll just use the Firefox. I can just start browsing, and every page that I look at through this tool is being recorded for me. So, I’ll start recording, initializing remote browser, and then you can just see this type of thing. So, now you can see every page is being saved to my web recorder.io account. So, I’ll look at this page, and I’ll look at this page, and now I can stop it, and I can go to my temporary collection. Obviously, if I have logged in and created a password, I have this in my own collection. What’s great about this is if you login, so now you can see that everything is saved. You probably can’t see at the bottom, but you can see the page is saved on recorder.io. And what I can also do this simply click and download this collection of pages for off-line viewing, and I now have a permanent copy of it on my computer. So, then I save these collections. I download another tool from web recorder.io, and now I can play all of those saved pages on my local system. So, I have a file I can look at on my own, or I can share with others, and now they have permanent access to these saved pages that were saved. For example, I’ll look at this page from the NFL.com that was saved on August 30 at 15:02 GMT. Pretty cool, huh? And, as of now, it is from a nonprofit, is from an arts organization, actually, and as I said, Mellon funding helped make it available.

In addition, we all know obviously about the WayBack Machine and their archive-it service, but there are other web archives out there. In the last couple of years, Herbert Van de Sompel at Los Alamos and others have created a protocol called Memento, and now they have an interface to it called “time travel.” So, again, I can now from one interface search multiple web archives, and while the WayBack Machine is great, it doesn’t have everything. This tool and another version of it uses the same protocol. It uses the same protocol, just a different interface to it in this case, from the UK Web archive allows me to search multiple archives with a single click. So, let me go back to Time Machine, and we’ll have it open up here, and we’ll just go to CNN.com. For example, I can say I want to find CNN from February 8, 2012, and now it will search multiple—so here’s one from the Stanford Web Archive created exactly on that date, different versions of it from archive-it, from the Alexandrian Web Archive, the Library of Congress. So, this tool allows you to search multiple web archives for pages using a single interface and also a free service, and you can do either from Time Travel, memento.com, or you could take a look at another interface. Matter of fact, with this interface, if I can get it to come up, there we go, there’s actually little bit—there’s a bookmark that you could move to your toolbar, and you don’t even need to type in the page. Whatever page is on your screen, it will run it up against the Memento databases.

**Rick Anderson:** Gary, do you know where the revenue is coming from to support this?
Gary Price: Memento? I think it is just a protocol that Herbert and others have created that allows the multiple web archives to talk to each other in a single search. So, it is a shared protocol.

Rick Anderson: Whenever I see something like that without advertising, I think, “Somebody is selling my web activity,” which personally I don’t really care about. I don’t do anything on the web that I’m embarrassed for people to know about, but it could be of legitimate concern for some people.

Gary Price: So, for example, here is again the University of Utah homepage. See if we can get Rick’s University, but, Maria, go ahead. Do you have any comments?

Maria Bonn: The web recorder, the thing I resonated to first and I wonder if some of you do as well, is what a great teaching tool! And I thought about sometimes inconvenience and sometimes perils of running live demos and live searches and say, “Here, let me show you how to do that.” Well, maybe you once can get it just right, and you have it stored so you can make use of it again, but also the sharing piece is very valuable. So, if you had something that you know your users are getting wrong time and time again, maybe because of your crummy interface, whatever, you can do it right and show them, “This is how you do this. This is how you find that out.” So, I was quite excited by that. We looked at Time Travel, and the thing that I spent last night doing was I ran one of the earlier library publishing operations. We started around 2000, a little earlier actually, and we changed our name and logo and URL at some point, and our earliest representation of ourselves is lost as far as I can tell unless it is buried somewhere deep in my personal archive as some HTML pages, and there is now a library publishing coalition. There are many members, and I was giving a talk, and I thought I should show them, “This is what we were thinking about in the first days.” Well, I couldn’t find it, right? So, I thought okay, this could help me solve my problem. It did, but I also thought and then I’m no longer at the University of Michigan Library where this happened. Once it’s turned up, Timewarner has helped me recover it, does it become University of Michigan library’s responsibility to archive this thing that is been found because we are a library? That’s what we do, right? Especially with our own history, or do we say, “Somebody is taking care of that.” That kind of goes back to the thing that Rick started with. How much should we own this activity and its product, and how much should we trust that, “Oh, we like Herbert; he’s a good guy. He’s taking care of it.”

Gary Price: I also wanted to mention real quickly, back to mentioning the WayBack Machine which we all know, one thing that often goes unnoticed though, and I like to do this to make sure that the copy or the report that I might be sharing with somebody or the document, the webpage, I know that the copy that I saw at that moment in time is available, and I might use web recorder, but another thing I could do and also share with the world is I could just take any URL of any webpage or any PDF that doesn’t have robots that text associated with it, drop it into the “save page now” box, and the WayBack Machine web crawler will come to it at that exact moment to make a copy of it and add it to the WayBack Machine.

To segue into another area is we’re seeing the camera phone become an information provider, and then, of course, image recognition, which brings up all sorts of privacy issues, but two tools that I will just mention very briefly, one is called Photomath. This is actually a point-and-click camera operation, a point-and-shoot operation. You don’t even have to be online. You point your camera phone camera at any math problem or algebra equation, and just by pointing at it, it will solve the equation or math problem instantly, and yes, Rick, right. This also gets those of us who didn’t do well in algebra or whose teachers kept saying, “We need to see your work.” This will actually show the steps used to solve the equation. As you can see, they just added the feature where it used to be that you had to have it in a printed book. As you can see, now they just added handwriting recognition to it. Getting more into image recognition and finding people and the like and actually showing products, this is what I have found to be one of the better apps for it. It works obviously with Android and with iOS, but basically, this is just point your camera at something, and it will use their database, their algorithm to try to determine what is and then run it against the search tool. I can say that I’ve actually pointed the camera at the television, and it has been able to recognize celebrities just by pointing the camera at the television and then runs the search against their database, so image recognition, as we all know, is getting better and better and better, and this is an example of that, really a tool that we could use and gives us people an idea of what is coming very quickly. That’s it for that section.
Gary Price: “Eewww!” for not making me scroll through a list of my friends,” and there’s part of me that goes, “Thank you for not making me scroll through a list of my friends,” and there’s part of me that goes, “Eewww!”

Rick Anderson: I guess I would just add that I have had mixed feelings about image recognition for a long time. Whenever I’m on Facebook, and I’m tagging a photo, and I mouse over one of my kids, Facebook says, “Oh, is this your son, Tucker?” You know, there’s part of me that goes, “Thank you for not making me scroll through a list of my friends,” and there’s part of me that goes, “Eewww!”

Gary Price: Right. They know his face.

Maria Bonn: About the math application, every parent in the room feels the same way, right? “Cool. I got a show this to my kid. No, I can’t show it!” And we were talking about this right before the session began that I have an 11-year-old who is deep into algebra and just getting introduced to trig, and I tried to help him with his homework, not always successfully. Well, this is going to solve a lot of problems, and then I had that immediate concern about people still need to know how to think. We need to know how to solve these problems, and I both laughed at myself thinking, “Okay, this is what my parents were saying when they first allowed us to bring a calculator into math.” “We used to have to do that with pencil and paper. A slide ruler.” So, I thought, “Don’t be old-fashioned, Maria,” but we want people to understand the principles at work behind what the machine does, and I suppose I was comforted by this showing of work piece of it. But, again, it left me thinking, “What should people still be doing and why do we say that? Should people still be doing this?” And the photo recognition app, I don’t think we scrolled all the way down, but at bottom, there is a thing about, “See what people are searching now,” and this may segue into our privacy conversation. Yes, what other Google searching around the world? And I was sort of speculating why do we care? What we want to know what other people are searching around the world? Is it that we’re a little insecure, and we think other people are always cooler or smarter than we are? They must be looking for something I should know about but that I haven’t thought to know about, or is it a desire to sort of know what is going on in the cultural/political/social/landscape? And this is a general question I have about services available to us online, what’s trending now? Why do we want to know so much with the crowd is doing? This is not at all to say that it’s not valuable, but I wonder about our reasons that we go in and we want to find these things.

Gar Price: In our remaining time, I want to point out that I have a bunch of reading things here for you regarding privacy, which could be another conference unto itself. I do have a couple tools. One thing I want to say about privacy is that everybody has got a different threshold and a different comfort level, but I really do believe that ethics issues, the fact that reader privacy is number three on the ALA ethics statement, and the same on other countries ethics statements and library organizations. All of that aside, I think the library community, we are always trying to find new ways to be relevant, and I can’t think of a topic that’s been in the mainstream news more than digital privacy in the last couple of years, and I think we could be serving our communities and become kind of a digital privacy clearinghouse. Share with people what’s going on, update them, provide them with the tools, whether that be having them come into the library, us reaching out, but I think people do trust librarians. We see that all the time, and I think informing the public about what’s going on and use it about what’s going on could be very, very valuable and very relevant to their lives.

One tool to help you learn more about digital privacy is called BrowserSpy, and this tool will go through all these different things and shows up to and will show you some of the different information and types of information that your browser is outputting that other people can see if they have the right tools. This is different than what I showed last year at the presentation when I had the Wi-Fi traffic going over. For those of you who were in the room, I was able to show the Wi-Fi traffic, all the unencrypted Wi-Fi traffic going out over the router in the center over there. This is other information that somebody could pick up from your browser, and it’s often being sent when you request a web page, everything from what the color capabilities of your browser, to your connections, of course, all the cookies.

One term that you might look at, and it’s in some papers there, which is a new form of tracking people on the Internet is fingerprinting, although they’re all different types of fingerprinting which can provide a specific person. It’s also been shown time after time, and this is different than it was years ago, but with enough anonymized data points, you can pretty much figure out who the person is and those bring up all types of surveillance issues as well. So, BrowserSpy is a tool that can be very informative in terms of learning more about what information you are sending out.
I've also given you links to a couple of tools that, again, depending on the user and their comfort level, can help block some of this information from going out. So, there is Ghostery, which is also a very useful educational tool to help you turn off JavaScript and other third-party trackers. There are companies out there that have cookies and other tools set up on web pages. These companies often come and go, and they provide little information, and they're tracking everything that you're doing online. So, there are a couple of tools out there that help block either all of it or selectively block some of these types of trackers and JavaScripts and the like. Privacy Badger, for example, comes from the EFF. There is a great page on the Google website, Google has more information about it than you can even imagine, but this is their privacy dashboard. One thing you can do here is look at what they're tracking and turn things on and off. For example, every time you do a search on Google, the default is turned on, but they're saving a copy of it specific to the IP address of where you're running that search from. They have that in perpetuity unless you turn it off or selectively go in there and remove material.

So, the Google privacy dashboard is a good place to begin. As I mentioned a moment ago, every time we search, there is an IP address that is sent out. There are tools out there that have maps, some better than others, depending on the location, but there are tools out there like for example this one, these GOIP databases. In some cases, they're down. For example, I was recently in Monterey, California. It actually didn’t even say, here we know that this IP address from this hotel is located in Charleston. These are these disparate pieces of information they can now be all merge together. In the case of when I was speaking in Monterey, California, it was not only that I was in Monterey, but I was actually in the Marriott Hotel in Monterey. All that information is going out there. With a little bit of knowledge, the end-user can determine if this is an issue for them and what to do about it.

And then I’ve given you a couple of other apps. Issues for the library specifically are what are we doing if we have cameras in the library? I’ve given you links to a couple of academic articles about that. We’re sharing a lot of data with Google, with Google Analytics. Some libraries are using it, and others are using the very robust open search package. Also, in the last year or so, the ALA has released privacy guidelines, reader privacy guidelines for the K-12 community. Are we going to see something similar in academics? And also privacy guidelines for e-book lending and digital contents vendors. But, I’d like to get my colleagues’ general thoughts in our remaining time about library privacy.

Maria Bonn: Oh, gosh. This is dangerous. When we regrouped a little bit yesterday to just say, “Okay, are we all on the same page for tomorrow,” we spent a long time talking about privacy, but one thing I was left thinking about was I was in academic libraries in the library for about 20 years, and I was involved in many really good soul-searching conversations about achieving the right balance between respecting or protecting patron privacy and building intelligent and informed, particularly Web services. They want recommender services. If they want it like Amazon, we have to know what they’re searching in the catalogs, so that means that keep that data. How long should we keep that data? These were all great conversations, and I think they should continue, but what I worry is that we have them and think, “Yeah, we are thinking about privacy,” when it’s also the library’s responsibility to educate our patrons so that they can make their own judgments about privacy, when to opt in, when to opt out, when do they need to take steps to protect themselves? And I’m anxious that we not abjure that responsibility because these things are gnarly and complicated. That’s exactly why we should embrace it.

Gary Price: Right.

Rick Anderson: Great point. I guess when I think about libraries and patron privacy, I find myself feeling really strongly about two kind of mirror image issues. One of them obviously is the library. We’ve got to be incredibly careful about how we manage the data about them that our patrons entrust to us. They do entrust some potentially sensitive data. The other thing that I feel like we have to be really careful about is not arrogating to ourselves choices that are our patrons to make about their privacy, and sometimes in conversation with colleagues, I will hear people say, “Well, we don’t offer that product because it requires patrons to give up too much of their privacy.” And I think “Dang, I’m just not sure that’s a choice that it’s your place to make for the patron.” And so I feel like these are genuinely difficult questions in which important principles are to some degree in tension with each other, and it’s something that I struggle with all the time in my own mind.
Who’s Faster, a Pirate or a Librarian?

Adam Chesler, Director, Global Sales, AIP Publishing
Scott Ahlberg, Chief Operations Officer, Reprints Desk
Carolyn Caffrey Gardner, Information Literacy Coordinator, California State University, Dominguez Hills
Georgios Papadopoulos, Founder and Chief Executive Officer, Atypon
Heather Wilson, Acquisitions & Electronic Resources Librarian, Caltech

The following is a transcription of a live presentation at the 2016 Charleston Conference.

Adam Chesler: Good morning, and welcome to the “Who’s Faster, a Pirate or a Librarian” session. I’m going to talk a little bit about something that’s been, actually it was addressed in just the previous session about Sci-Hub, about the ways people get access to content, the perhaps questionable legality of some of that, but there are reasons why it happens, and we want to try and address some of those things today. We have four speakers, oh, it would be appropriate to tell you who I am. I’m Adam Chesler. I am with the American Institute of Physics Publishing, and I’m here to simply guide us through this. But, our speakers today, and their biographies are available to you, so I’m not going to read them all to you, but the order in which they are speaking is Carolyn Caffrey Gardner, from Cal State University in Dominguez Hills. We have Heather Wilson from Caltech, Georgios Papadopoulos from Atypon, and Scott Ahlberg from Reprints Desk. What we’re going to try to do today is look at very broadly some of the legal, practical, financial, and technical issues surrounding access to content and the ways people are trying to address this today as opposed to, as Chuck alluded to, 20 or 30 years ago. So, without further ado, I will let Carolyn get started, and we do expect to leave time for questions, so I’ll simply ask that you hold them until the end so that we can cycle through everybody’s presentation as best as we can.

Carolyn Caffrey Gardner: So, thank you for having me. Just a little bit about myself, I’m an Information Literacy Coordinator, so I’m really approaching this problem, this issue from the point of view of our users and sort of really like, “Why?” And how can we really look at instruction in ways to really challenge with happening? So, I first got interested in this project when I was on my own personal Twitter timeline, and I noticed that there were these tweets with the #icanhazpdf. This is just a selection of a few that I saw, and so I was very curious about this and tried to do a little bit more research and try to figure out what is #icanhazpdf? What are these people doing? And found that there wasn’t a lot out there, and so I really started exploring it. Within the tweets we noticing all sorts of things like people saying, “I don’t have university access.” “I can’t get this through interlibrary loan.” I found this one particular tweet: “I can’t find my own paper. I don’t have access to my own research.” And so I was very curious and kind of exploring what’s happening here? What’s being requested? And that is continued on into Sci-Hub and LibGen.

So, in terms of resource sharing outside of libraries, there’s really sort of two things that are happening. There’s peer-to-peer networks like we just saw with #icanhazpdf. It’s a one-to-one person exchange. Reddit Scholar has a very robust community. It’s got a couple hundred thousand people in it, very similar process. People post a DOI, a link to an article, and an e-mail address. It’s posted; they remove their link and their information there. There are closed Facebook groups, and then this is not new. Right? It has been happening through e-mail, photocopying articles for colleagues. There’s other peer-to-peer ways. It’s a little bit cut off at the bottom but there’s also these larger repositories. So, we have got Sci-Hub, LibGen, and Avaxhome. LibGen has been around since around 2007. It’s one of the big repositories that powers Sci-Hub. There was a study in 2015, I’m probably pronouncing that name wrong. Cabanac? And he really looked at kind of what is in LibGen and found that for the top three publishers, Elsevier, Wiley, and Springer, that 68% of their content was in LibGen. So, without LibGen, Sci-Hub would not exist.

And then I’m also curious who is using these sites? This is pre-John Bohannon science articles, so I did not have a lot of raw data, but I used Alexa Internet...
traffic data to really kind of look at who are visiting these particular sites. So, Sci-Hub since the last time I did this was a year ago, and it has gone up about 19,000 places in global rank, so it is growing in popularity. The United States and Brazil are only now just making the top five countries in terms of use, and they kicked off Russia and Indonesia, but I point this out to show that it’s not particularly just a U.S. phenomenon, even those that may be the context from which we are approaching it. And this, of course, does not include mirrors or hard-drive version of Sci-Hub. I was looking at SciHub.bs for this, but there is also a SciHub.cc, .io, different mirrors.

And for LibGen, which powers Sci-Hub, the global rank is significantly higher, and again, the United States and Brazil have moved up kicking out Indonesia and Russia from the top five over the past few years, and I point this out to take a look at Iran in particular there and the rank and country, so it’s in the top 1,000 websites for Iran, so kind of let you ponder, “What does that mean?”

In terms of my methods, and my other co-author, Gabriel Gardner, could not be here today, but we really did two things: We archived a bunch of tweets of #icanhazpdf to see what are people requesting? And then, our later project, we surveyed users of these peer-to-peer sites as well as these larger repositories. It was a convenient sample. We're getting people who are using these particular materials. Many of them reached out with us with very long e-mails and voicemails and then also Bohannon’s Science piece. He was kind enough to give us the raw data as well, so we were able to take a closer look at that.

In terms of who are people using these services, what is being requested? Important there. When it came to Twitter, it was primarily journal articles. One of the things that I found really interesting is there was really no one journal title or one publisher represented. There were 494 unique journal titles within the 674 unique requests that we analyzed, and I thought that was really surprising, so it wasn’t something about a particular “big deal” or a particular expensive journal that people didn’t have access to. It really was across all disciplines, all publishers. Likewise it wasn’t just new content that might be embargoed. Only 30% of the requests were published within the last year, though life sciences, biomedical sciences definitely had a larger percentage. Arts and humanities was a very tiny sliver, but they are there. Again, in terms of Sci-Hub, the top three publishers downloaded were Elsevier, Springer, and IEEE, again, heavy representation from the sciences.

So, who are the people doing all of this? Within our survey we asked them, “Are you affiliated with a university? Do you have access to journal articles?” And again, surprisingly, only 20% said they were not affiliated in any way with a university. I thought this would be much larger. I’m not sure the 6% who are not affiliated but also not any part of a university what their affiliation status is? Grad students did make up a large portion of who is using these services. When we asked them the frequency with which they use the services, grad students use them more than once a week, and people who used LibGen and Sci-Hub were more likely to use them more often. #icanhazpdf and Reddit Scholar was very much like, “Once a month when I can’t get an article,” type deal.

We asked them why? What are your motivations for obtaining materials this way? And it was an open-ended question so we then coded the responses based on different themes that we saw. Not surprisingly lack of access was the number one reason why people said that they use the services. There were a lot of people who wrote on our survey, “Why are you asking me this? Of course, I don’t have access. That is why I’m using it this way.” And speed was another one. There were a lot of comments around interlibrary loan and things being slow, which is one of the reasons we got talking is speed was such a pervasive issue. It wasn’t just access. Within that, we saw some comments about user experience, some about cost, and, of course, there were some people who had ideological motives. There was a lot about open science and free science.

Thankfully, this matches up with John Bohannon’s survey that he included with his article about Sci-Hub and science, so we’re not totally off base with our smaller sample size. Within his large sample size, this is cross-tabulated data on two questions: Have you obtained a pirated journal article through Sci-Hub or other means despite having access to it? And what is the primary reason you use Sci-Hub? So, you can see, among people who have access and people who don’t, you’ve got under no access that is still the primary reason. For people who have access, people who don’t, it was their primary motivation. But convenience is in there as well, right?
We then kind of drilled down further. People who said they use interlibrary loan, we asked them why aren’t you using interlibrary loan here? What determines whether you obtained materials through interlibrary loan or another means? And you can see some of my favorite responses are here. There was a lot of, “Slow, it takes forever.” “Interlibrary loan isn’t free at every institution for users.” And, you know, the $5 charge might be too much. I liked the last one. That happens to me all the time too.

We asked them what do you think about potential copyright or terms of service violations within these different peer-to-peer services but also these larger repositories? There are a lot of “don’t care,” whole pages of “don’t care,” “don’t care.” These are some of the more robust responses. In here, when we asked people what they thought about copyright and terms of service, we saw a lot of responses related to ethics and ideology. You know, you can see here this person says, “Data should be free.” So, why should we care about this? Well, it’s probably not going away. Crowdsourced communities are motivated by these sharing and reciprocity goals. When our data we asked people how often do you post articles? How often do you get articles? And there were very few people who are just leeching off of the system. They really are communities doing both, very similar to file sharing of Napster in days of yore. These systems are decentralized. They’re largely pseudonymous, so there is a low likelihood of punishment, though I’ve gotten in some hot water for saying this previously, they are very easy to use. Sci-Hub has a chatbot. You put in a DOI. You automatically get a PDF right there. There is no click-through. Again, Association of American Publishers, don’t write me another letter. I am not advocating the use of the services. However, if you did want to look at the usability, there is open access content within these things as well, so you can look at one of those open access journals. And again, they have multiple mirrors, so not represented within a lot of this data is the fact that folks in Iran and China in particular download and put on external hard drives a lot of the data from LibGen and Sci-Hub.

So, closing thoughts. Again, I’m an instruction librarian, so I think one way to address this problem is greater focus and information literacy instruction, particularly on the information has value part of the framework. So, how do we work to educate users? There were a lot of comments within Twitter about how people losing access to library resources when they graduated and not fully understanding what that was all about because all this time we as librarians have been saying, “It’s free!” And not really putting in the, “... while you’re here.” Or, “It’s not really actually free. We are paying it for you.” So, we can do a better job there. And then I also think this is a social justice issue as well. It is not a U.S. phenomenon. How do we approach this from a global perspective? References are there, the full study as a preprint and college research library. And again, thanks to my co-author who couldn’t be here today. So, without further ado, I’m going to turn it over to Heather.

Heather Wilson: Thanks, Carolyn. That was so interesting. So, I’m Heather Wilson. I’m an acquisitions and electronic resources librarian, so I deal primarily with mostly the linking and technical pieces of this, and really even in my acquisitions role, it’s largely a technical role as to how those acquisitions are being accessed. So, I want to talk about some reasons people might be having those access issues that are not financial or they’re not maybe even related to whether or not the library has the PDF to begin with. Yeah, because there are a number of reasons people may want to use Sci-Hub. Most of this will probably not be news to this room and just a resonant. So, some of the things that I want to talk about, first, okay, our linking difficulties might be a major reason that people use Sci-Hub. Of course, open URL failures are a big thing. Generally, people working with having their searches optimized and having their searches made as ideal as possible where we are not necessarily doing so and then large and complicated networks. This is like my favorite thing to cite all the time, ever since I learned about it. I think Ken Varnum is probably here at Charleston, but the University of Michigan did a study on their open URL resolver that I found very interesting. They wanted to see how successful that linking was happening, and just to give you an idea they started with, I want to say it was in the hundreds, but they started with a sample where they have self-reporting. They had an option for users to self-report when they ran into an issue very quickly, and then they also picked a sample of articles to use for their own testing and linking. Just to give an idea, this is the direct linking. This is kind of the traditional idea, and you can see the trends over different periods where they measured, and it’s pretty much what you would expect, you know around 97%, 94%, barring technical difficulties. But, then they studied it through their open URL resolver, and the numbers came out
significant. They found consistently that, yeah, they were only successful linking 63% of the time in the most recent sample. So, that’s 40% of articles that are not being linked to through the open URL resolver, and Sci-Hub does not have to deal with this, and so that might be a very good technical reason why somebody might go to a central place.

Another thing is the optimization of searching that can happen at one level. People are used to searches that have been SEO’d, used to searches that have been optimized to make their “one click” experience as good as possible and things like they’re focusing on targeting “long-tail” keywords, or as we know them “known item searching,” people who search for very specific items, whereas we seem to often be thinking more broadly about concepts. We seem to often be going in a different way. Thinking about improving their “bounce rate” and reducing the number of clicks and keeping people on the page longer, and which sounds like something that we would really be concerned with, and so, yeah, I think the average open URL linking goes through at least three clicks, and I think there is an average of five on some repositories and other types of open access things. And so as a result, yeah, we’re certainly not meeting their experience. And the other thing is, of course, large networks. This is the publishing cycle, and every place that has, and you guys are all familiar with this. Every place that has a stop along the way, but it is, of course, important to know that each of these things have different servers, each of these things have different authentication measures. I’m gonna move real fast.

Another reason might be indexing and discovery insufficiencies. There’s a lot of stuff that is just not being exposed. At Cal Tech, we are very concerned about hybrid OA, but there’s also massive metadata failures, and the knowledge base structures don’t always meet the content and the arrangement that it. So, one example, the hybrid OA situation, so one of the problems that we are having is that we may not have access to a journal, or we may not build to subscribe to a journal, but we have researchers who have published open access in that journal, but because we can’t list it as one of our holdings, we can’t really expose that content, and so that hybrid OA, the open access that is within a journal is very difficult for us to list. Another situation related to that would be like green OA, which doesn’t really often have a lot of the structures that are required for open URL. It may not have a DOI. It may not have a volume issue number, and Sci-Hub is acting as a search engine, and so, therefore, it doesn’t have to be concerned with those structures.

A metadata insufficiency, of course. This is another thing I like to quote every chance I get from Kristen Wilson of the GOKb Project at NC State, and the metadata is not always what it should be because there’s not really a true caretaker. As she says, “Publishers are in the business of selling content, not metadata.” They’re working to meet the user experience as well in the end and not necessarily as concerned with getting us all the things that we need to make that linking work, and that is a huge reason that we might have issues that Sci-Hub doesn’t necessarily have to worry about. And, of course, knowledge base structures are pretty huge. Right now, I think most people in here have to work with a knowledge base that works at the publisher level. You list your titles under the publisher. You list your articles under that, and there’s a certain hierarchy here, but if something doesn’t have a publisher yet, such as the case with green OA, or it is a preprint or may not be linked to a title or increasingly a repository items, and there’s not a good structure for that in the knowledge base. You kind of have to force it in there.

And then authentication barriers are, of course, the biggest ones that are probably the most common so I will go really fast through these. So, of course, DRM restrictions. We hear it time and time again. You buy the item. You download it properly. You lose it, and you end up a criminal trying to get it back where you downloaded it the first time, and you’re a criminal. I think this is probably the most cited XKCD comic ever, but this is a major issue, and, of course, Sci-Hub doesn’t have to worry about it that as long as we have PDFs and e-books with limited containers, we imagine that this is gonna be a problem. And then, of course, user privacy ambiguity is something that is rising. I include this graph just to show—this is from a Pew Research Center study on privacy where they asked people the different extent to where they’re concerned about their privacy, but I mainly just include it to show there is a rise, a growing trend where people are concerned about their privacy. So, things where people have to create logins, things where people are very aware that they’re giving personal information seems to be on the rise, and more over the idea that information has value. So, even people who aren’t necessarily concerned about their rights are suddenly aware that they have a
commodity thanks to the global conversation, which I’m thrilled about, but as a result, that is something that becomes less problematic when you have fewer logins and pseudonymous operations. Alright. So, those are my references, and this is me. Thank you.

Georgios Papadopoulos: Hello. My name is Georgios Papadopoulos, and I am the CEO of Atypon. Let me tell you a few things about Atypon. Atypon is a technology company in the business of delivery, of content delivery building the websites and delivering the content for a number of publishers. We serve about 10,000 articles and about 40% of all research content. So, we work exclusively for publishers, and we try to provide them technical solutions to any issues that they have. We don’t make any judgment as to the business models or anything else or the social issues. These are between publishers, regulatory authorities, and libraries to solve. We’re there just to advise on a technical level.

So, Sci-Hub, of course, has been a big issue for publishers. Many of them don’t know some of the other places. They know LibGen and some of the other places where the content is leaking, and, of course, that is a problem for them because it threatens to destroy the whole ecosystem of publishing. From a technology point of view, however, I really want to thank the Pirates in this case. It’s been 20 years—I started a company with the first journal that went online, the Journal of Biological Chemistry, and really it has been a struggle to make publishers, and I think to some extent librarians, move over from these standards that were established back in ‘95, the proto-web as I call it, standards which were the ID authentication and a little later, the PDF didn’t exist in ‘95. I think it came out in ‘97 or ‘98. I don’t remember exactly. And PDF, and it’s been a struggle because everybody has really acknowledged all these 20 years that these were actually bad standards. They’re not really serving the users well. They don’t provide the right user experience; however, nobody wanted to change them. So, this is the way, the reason I actually thank the Pirates because they’re forcing the change that the technology companies could not actually force the publishers to do.

So, let’s start with the big one: IP authentication. IP authentication, everybody has acknowledged that it doesn’t actually identify the institution very well, that it creates problems for the librarians. They always have to update their IPs on over 50 or 100 or over 200 sites. I don’t know how many sites anymore, and, of course, there are all the problems with remote off-campus access or institutions that don’t even have stable IPs. All kinds of things are there that you can find once you go into that, and there is really no reason for that. The technology for us to move over from IP authentication and create this frictionless experience that Sci-Hub has where you actually hit a DOI, and there you go to the content that you are—and, of course, Sci-Hub doesn’t care about entitlements, but we assume entitled content. For me, as a technologist, the biggest problem is that many, many users actually have access to the content, but they are so confused with all the rules and all the things that they have to do that they actually lose access to the content. So, for me, it is very important that a user logs in on his device once in his lifetime from his institution, and from then on, okay, he logs in without even being asked to any publisher what is his IP, what is his username. He’s not handed any tokens, codes, whatever, that he has to put into his device every time he starts a session, and this is something that we’ve actually demonstrated right now to the publishers, and the project is called I think Universal Researcher Access. And hopefully you’re going to see it rolled out in 2017, so, of course, once you don’t have IP authentication, there’s no more Sci-Hub because Sci-Hub really depends on having IP authenticated access through institutions. Okay?

So, let’s go to next one. Since we are in authentication, I thought let’s throw passwords into it as well. I mean, okay, I understand that we have used usernames and passwords for a long, long time, and we’re getting used to them by now. Of course, we are always hearing in the news about compromised sites, about stolen passwords, about what you don’t know, of course, is all the password tracking attempts, fishing that goes on, identity theft and all these things. There’s really no reason to have passwords. We can achieve perfect access without any passwords. If any of you have used medium.com, it shows you why passwords are actually not necessary. We don’t need them, and if we get rid of passwords, we get rid of all kinds of problems associated with passwords. We don’t even need to remember them too, which is a good thing.

Next one, the PDF. That’s a big one. Everybody has come to love PDFs. I don’t know why. I mean, they’re really there to be an electronic equivalent of the print, and in the time that we don’t print
anymore, why do we need this? I don’t know. It’s really—we create this artificial thing where even journals that are not even printed at all—they have PDFs. Amazing! How do they come up? So, and meanwhile they have all these problems, so if you cannot deep link the PDF, try to read the PDF on your smartphone. It’s not referable. It’s really terrible in terms of user experience, yet I can tell you users 3 to 1 or 4 to 1 use PDF frankly because the HTML that the publishers produce. It’s not the HTML; the HTML is good. The HTML pages that the publishers produce have so much crap into them that nobody actually wants to read them. So, anyway, there is a portable format that has been around for a number of years, and it’s open standards, and it’s called ePub. It solves really all the problems that PDF has. It has all the deep linking, and it’s referable, and you can view it in a browser, and it even has an open standard for DRM, so if somebody wants to enforce DRM so that only the people who have access to that ePub can read it, then it’s possible to do that. So, of course, once we move from PDF to ePub, guess what goes away? #icanhazpdf goes away. LibGen goes away. All of the other stores where you have unauthorized posting of article, all of these go away. So, it’s going to improve the user experience, and it’s going to let the publishers and the libraries work out all the business models that they want to have, and that is what the technology can do. Thank you.

Scott Ahlberg: All right. I thought I would start off my section with a haiku. I’ve worked in a few different information companies, and about 10 or 15 years ago, a company that I was at to try and liven up the workplace decided they would have a haiku contest. And it had to focus on the work we were doing, and this is one of the entrants that kind of stuck in my mind and seems somehow relevant to this current discussion:

Copyright, piracy
Information wants to be free
But no, it is not

As Adam has mentioned in the introduction, we wanted to take a look at this from multiple angles, the challenge of piracy and what it means for what we all do. So, I think there are multiple meanings in this haiku, and I think with the current situation maybe even a couple of new meanings have come in. I attended Carolyn’s presentation at ALA in Orlando in July and was really struck by the statistics that she gave, particularly in the motivations for using Sci-Hub which she gave a few minutes ago and the speed access of it. It’s—notice that cost is down there at 13%, whereas speed is at 26%, so I think the question that I would hope that everybody in this room asks himself is how you want your user community accessing the content that is available, that maybe you’ve already paid for or that you would be willing to find some way to pay for if the use could be granted to them? Now since I’ve been in the information business for a couple of decades now, I always find it informative to take a look at how other content industries deal with issues that are related to their delivery of their forms of content. So, I often take a look at what’s going on with the delivery of video and what are the business models that are in use there, delivery of audio. I think software is different enough that it is maybe not quite as informative, but I think particularly with music delivery, obviously the Internet has brought us all kinds of forms of disruption to all kinds of business models. I think we all know what’s happened to the music industry, so I think the lessons that I see in the music industry is that initially with Napster being sort of forefront of disrupting the music industry as we knew it previously. My interpretation is there was a real failure to pay attention to users and what users wanted, and while I think that that’s been turned around somewhat very recently, and I think, I’ll speak for myself here, as a listener of music, I’m much more satisfied with the options that I have available to me today than I was 10 years ago. I think that there’s been lasting damage in the music industry by a failure to pay attention to what users want and need. I think the evolution in the video industry is a little different, and I think that there’s been a higher level of success, a higher level of user satisfaction perhaps, and I think less of a lasting damage looking at it from the perspective of the content producers and content owners, less of a lasting damage to the business model.

Now, of course, copyright law underlies a lot of this, and I think again if you go back to the pre-Internet era, copyright law at least in the US was a matter of civil law, but through disruption we’ve seen changes to that. I think most noticeably the Digital Millennium Copyright Act has criminalized some aspects of copyright law, and that certainly comes into play any time you have DRM on content and then maybe a step or two removed, but the Computer Fraud and Abuse Act is something that
has been looked at as a potential way of addressing copyright infringement if it involves sharing of passwords or breaking what would be viewed as appropriate authentication to access that content.

So, when I started looking at putting this presentation together, my original intent was to give a little bit of an overview of the legal landscape and whether any of the legal challenges that have happened in other content industries might provide any sort of pointers or guidance or foreshadowing, however you want to look at it, to what might happen to the scholarly publishing industry and the user community if piracy continues in the way it has been. Are the users at risk? Are libraries at legal risk? But, not being an attorney, I wasn’t prepared to give any sort of a legal analysis here, and I ended up changing my presentation just a little bit after attending the green and gold open access session on Thursday afternoon. I don’t know if Jason Price is in the room? All right. Excellent. Well, I thought I was an excellent session, and I really enjoyed the way Jason presented the overview of open access and piracy and the user experience and what’s available and while I don’t think that it’s—I wouldn’t otherwise want to include open access and piracy in the same discussion, I think what really came out from Jason’s presentation is that, from a user perspective, it doesn’t matter whether it’s pirated content or open access content. They just know that they want the content, and the distinction of whether their access is legal or not is often not necessarily known to them, or they don’t necessarily care. So, certainly I think one of the differences with accessing scholarly publications, and I think in the last session there was quite a discussion of fair use. I think the fair use aspect makes it a lot less clear whether the—if we’re talking about scholars accessing content, at what point does the fact of whether that content was pirated or not, at what point does it matter? The point is that they want it quick, and they want access, and I don’t think—Carolyn makes the point that it’s a social justice issue. I don’t think it’s the intent of anybody to deny access. We just need—there’s a business model behind publishing that obviously needs to be supported somehow.

I think one of the other key differences here in looking at the scholarly publishing industry and user community in contrast to music and video, for example, is that with scholarly publishing there is a professional class who is responsible for curating and ensuring access to the content. Well, obviously that doesn’t exist in music and video. So, I really appreciated Georgios sort of laying out some of the basic elements of the solution. In the time that I’ve been looking at this challenge, I don’t myself have any—I don’t proclaim to know what the solution is to the piracy challenge. I do think, however, that we need to pay attention to the users. I think that was the point that I really got out of Jason’s presentation. I think that’s the lesson that we can learn from the music industry is that it’s essential to pay attention to what users are doing and what users need to do and want to do and what their work habits are and what their workflow is in any solution that we come up with. Otherwise, we’re not actually solving anything. And, I think if we fail to do that, the challenge of piracy is an existential threat, not just to the publishing industry but to libraries as we know it because I think that is as we heard both from Heather and Carolyn that users are essentially going around the library in order to get the access, whether it is easier, faster, whatever it may be. Thanks.
You Can’t Preserve What You Don’t Have—Or Can You? Libraries as Infrastructure for Perpetual Access to Intellectual Output

Anja Smit, University Library Utrecht

Abstract

Since their existence, libraries have been responsible for preserving society’s records and intellectual output. This ancient and important role is under serious threat in the digital age. Even for scholarly journals, the issue of perpetual access has not been solved, other than by libraries buying access to archival materials. Recently, it became clear that the open access business model, with a focus on free access to new publications, introduces new problems for the archival role of libraries.

If ownership is crucial for preservation, who will ensure future generations have access to scholarly journal content of past times? Who are the actors in this new environment, and what opportunities can be identified to address this important issue?

Libraries have changed immensely since ancient times. At the same time, libraries have not changed at all. Sure, I could see how a Roman citizen would not necessarily recognize the Hunt Library as a library. However, some things about libraries have not changed in over 2,000 years. Today, I would like to focus on the core mission of libraries: To ensure perpetual access to knowledge (mostly documented, written text).

This was the mission of the libraries in Efeze and Alexandria, and it is still our mission today. It may not be the only thing we do, and we might go about it very differently today, but it is still a very important part of our added value to society.

The digital network is the perfect environment to advance access to knowledge. Naturally, we were there from the start of the digital era to identify opportunities and provide better services to our users. Of course, we embrace new technologies to fulfill our mission and add value to society.

At the same time, we should also acknowledge that in some areas networked services of new players are able to replace library services, so our services need to evolve. We need to review critically where we really add value and where this might be less so.

Two years ago, my colleague Coen Wilders addressed this conference on the topic of local discovery. It is our vision in Utrecht that we want to serve our users close to where they are on the digital network and provide access to knowledge within their workflow as seamlessly as possible.

Here are some of these workflows (referring to slide). Over 40% of the traffic to these publishers’ content comes through Google and Google Scholar, and we know from another big publisher that this is over 50% both for books and journals.

This is what researchers themselves say. Over 20,000 researchers responded to a survey from two of my colleagues, Bianca Kramer and Jeroen Bosman, on the use of tools on the Internet. You can clearly see how small the role of the library is. Our strategy at Utrecht is to invest less in local discovery services. Rather, we focus on the delivery of the content through the systems our users prefer, for example, Google Scholar.

While libraries may be less important for users to discover knowledge, they are still important to provide access. However, with open access advancing as a publication model, Google digitizing the world’s books, and users organizing their own access, the role of libraries for delivery may become less important too.

Will open access become the publishing model of the future? Some believe it will. In the United Kingdom and the Netherlands, golden open access is on the agenda of our national governments and university administrators. Since 2015, in the Netherlands, license negotiations have included demands for open access for publications of Dutch researchers. We focused on eight large publishers. So far, we were successful in six contracts. In these licenses, our communities have access to the content as before, and the publications of Dutch researchers are published immediately in open access by the publishers.

In two licenses, the cost model was flipped: We now pay for the publishing services instead of for reading.
rights. This results in a situation where access to current Dutch research is open to the world. Wonderful.

However, in some cases, the publisher negotiated temporary access instead of perpetual access to the content. When I was personally confronted with this during license negotiations last year, it proved difficult to effectively make the case for perpetual access. How to make the case for eternity? To be honest, I may not have done a very good job, for the result was not satisfactory. The contract to be signed will lack perpetual access rights to the content, starting in 2016. Even worse, we do not know for how long this content will preserved. What we do know is that if publishers are the only ones to be responsible, long-term preservation will depend on commercial interest, and this is just not good enough. No research can be done without access to knowledge, and this includes insights of previous times.

Only a selection can be preserved, not everything. This was always true, and it is still true in the digital era, but perpetual access should be organized by trustworthy organizations that use appropriate criteria that are important for research and ultimately, society. Of course, this is not new. Libraries and other organizations in the public domain have been aware of the need for preservation of digital content from the start of the digital era, including post-cancellation access to commercial content. Many organizations did invest in long-term preservation.

Solutions available today so far include:

- Dark archives.
- Pre-print repositories.
- Archives on CD.
- Archival rights based on goodwill.
- Unclear post-cancellation agreements.
- Reports on evolving collections.
- First steps to change copyright laws to enable archiving.
- Collectively address the issue of rolling back files policies of publishers.

There are promising initiatives too. Several national libraries have already built repositories, including commercial content. Some organizations are creating infrastructures for open knowledge. They do so with or without other partners. The National Library in my own country, the Netherlands, has been building an e-repository since the 1990s, but this is a dark archive. The French National Digital Library, a project by INIST, is building an archive for long-term preservation including access. It takes a Frenchman to think of keeping content within the national borders, of course, but it is still very useful. There are many more such examples.

There are two questions, though:

1. **How do we scale preservation services?**
   
   In many cases, initiatives are national projects. Is this sufficient in a world where research is increasingly carried out in international networks?

2. **Who are important stakeholders?**
   
   Can libraries do this by themselves? Who should they partner with?

In terms of scale, four or five copies of knowledge hubs should be enough. More realistically, a network of preservation hubs might consist of local solutions, connected through standard protocols, but whatever way we do it, if we want to serve our research community, we better collaborate and create global access to global knowledge, content that is preserved and available to next generations.

Who should act on this?

First, libraries do not own digital content but can collect it. I must say I never believed in libraries collecting the institution’s research output, but it may become necessary. Theoretically, if we all do that, we would effectively be creating this knowledge hub.

The case of open access shows us that if we aspire to make scholarly communication more effective we have to get the business case on the agenda of other stakeholders. After a decade of libraries promoting open access to publications, things only started to change when governments, policymakers and administrators made it a priority on their agenda and when it became part of their business case.

For example, when in my part of the world open access, open data, and now open science became part of the agenda of the European Union, things moved along. Copyright issues are debated, and infrastructures are built.
Here’s I think what we should do:

1. Make perpetual access to knowledge the top priority on our agenda. It is at the core of our mission, and society runs a risk. If we don’t do it, nobody will.

2. Get perpetual access to knowledge on the agenda of relevant stakeholders as quickly as possible and do it collectively. Certainly, in a world where important stakeholders see the need for open science, we have a huge opportunity.

3. Find partners to develop long-term preservation infrastructures.

Let’s contribute to the availability of knowledge to future generations and leave the rest to Google.

Thank you.
The Long Arm of the Law

Bill Hannay, Partner, Schiff Hardin LLP

Abstract

This presentation provides updates on three legal issues of pertinence to librarians: The “right to be forgotten,” the application of the Americans with Disabilities Act (ADA) to libraries and universities, and the application of copyright “fair use” doctrine to electronic reserves and electronic course packets.

An Update on the “Right to Be Forgotten”

As you may recall from prior “Long Arm of the Law” presentations, the European Union vigorously protects privacy rights. Twenty years ago, the European Parliament and the Council of Europe adopted the EU Data Protection Directive, that is, Directive 95/46/EC of 24 October 1995. It protects individuals with regard to the processing of personal data and the movement of such data.

What is personal data, you may ask? It is any information relating to an individual, whether it relates to his or her private, professional, or public life. It can be anything from a name, a photo, an e-mail address, and bank details to posts on social networking websites, medical information, or a computer IP address.

Two years ago, the European Court of Justice had down a landmark ruling in May 2014 that EU privacy law required Google to take down (or “de-index”) negative information about an individual citizen of Spain, Sr. Mario Costeja. See Google v. Agencia Española de Protección de Datos, Case C-131/12.

On May 13, 2014, the ECJ held that Google (as an operator of a search engine) is obliged to remove from the list of search results any web page links relating to an individual if such information is irrelevant in relation to the purposes for which the data was collected or processed and in the light of the time that has elapsed. In short, the ECJ required a balancing of the legitimate interest in access to information and the data subject’s fundamental rights.

The court’s decision opened a floodgate of privacy requests from other EU residents. In the past two years, Google has received a half million requests to remove information and has complied with 43.2% of them. While many applaud this development, there has been some fear among historians and librarians that the role of libraries in preserving historical records is being impaired.

The 1995 EU Data Protection Directive will be replaced in 2018 by the General Data Protection Regulation, but the new rule will not cut back on the right to be forgotten. EU citizens will still be able to request data custodians such as Google to remove negative information about individuals, but there remain limits on it, as Viviane Reding, Vice President of the European Commission and EU Justice Commissioner has remarked:

The right to be forgotten is . . . not an absolute right. There are cases where there is a legitimate reason to keep data in a database. The archives of a newspaper are a good example. It is clear that the right to be forgotten cannot amount to a right to re-write or erase history. Neither must the right to be forgotten take precedence over freedom of expression or freedom of the media.

The latest controversy about the right to be forgotten is the ruling of the French data protection agency (CNIL) in September 21, 2015, now on appeal to the French courts. There, the CNIL ruled that Google must take down or delist results on all of its extensions, including its U.S. portal, Google.com. The ruling is not just limited to Google’s European ones. Thus, the French ruling would directly affect searches done in the United States.

The International Federation of Library Associations and Institutions (IFLA) is a strong voice urging restraint in applying this privacy right. Most recently, in an October 2016 letter, IFLA urged the French courts to reverse the state agency and not to expand the right beyond national borders.
Can the ADA Spell the End of MOOCs?

On August 30, 2016, the U.S. Department of Justice (DOJ) formally notified the University of California at Berkeley that it had violated Title II of the Americans with Disabilities Act (ADA) by making free audio and video content available to the public on YouTube and iTunes and in massive open online courses (MOOCs) but not making that content accessible to the deaf and the blind. The DOJ advised Berkeley that it must modify its free offerings and pay compensatory damages to aggrieved individuals.

In September, Berkeley issued a statement that it is, in effect, between a governmental rock and a fiscal hard place, unable to afford the cost of restructuring the programs. It may, therefore, have to remove the content from the public. Sadly, this is a no-win situation.

Berkeley is not alone among schools that have been sued by the DOJ for ADA accessibility violations: 25 others have too.

Where will it all end? It is hard to say at this point. Perhaps the Trump administration will take a different view of the situation.

Georgia State—e-Reserve Case

As you may recall, Georgia State University became the target of a copyright suit for allowing professors to designate portions of books and periodicals to be copied by the library, scanned, and put on electronic reserve or compiled into electronic course packets. Three publishers (Cambridge University, Oxford University, and Sage Publications) sued, alleged that substantial portions of 6,700 works had illegally been copied and transmitted to students for some 600 courses at the school.

After discovery, the case proceeded to trial, and in 2012, the district court largely ruled for Georgia State, holding that it was “fair use” for the university to electronically copy up to 10% of a book or even a whole chapter. Georgia State University v. Becker, 863 F. Supp. 2d 1190 (N.D. Ga. 2012) (Evans, J.).

In 2014, the U.S. Court of Appeals in Atlanta reversed and ordered the trial judge to take another look, using a more nuanced analysis. Cambridge Univ. Press v. Patton, 769 F.2d 1232 (11th Cir. 2014). Significantly, the appeals court held that the nonprofit, educational nature of the university’s use of the material favored a fair use finding.

Publishers were horrified. They looked at this sort of wholesale copying as undercutting the entire ecosystem of academic publishing. They hoped for a better result on remand, but that did not work out for them. In March 2016, the trial court again ruled in favor of Georgia State after taking a second look. The court largely tracked the same logic as before.

Where will it all end? Spurred by the apparent success of Georgia State, other colleges and universities have adopted similar eReserve and/or eCoursepacket approaches. Publishers have fought back, filing similar cases against U.S. universities, including UCLA, and against foreign institutions, including York University, Delhi University, and in New Zealand. The jury is still out, but the publishers have so far not done well in the Indian case.

Delhi University Photocopying Case

In September, a trial court in India ruled against publishers in an even more blatant case of copying, one where the university worked directly with a photocopy service to make hardcopy course packets for sale to students. See University of Oxford et al. v. Rameshwari Photocopy Services et al., CS(OS) No. 2439/2012, High Court of Delhi, Decision dated 16 September 2016. The trial judge stated:

That, in my view, by no stretch of imagination, can make the [photocopy shop] a competitor of the [publishers]. Imparting of education by the defendant . . . University is heavily subsidized with the students still being charged tuition fee only of Rs. 400 to 1,200/- per month. The students can never be expected to buy all the books, different portions whereof are prescribed as suggested reading and can never be said to be the potential customers of the plaintiffs. If the facility of photocopying were to be not available, they would instead of sitting in the comforts of their respective homes and reading from the photocopies would be spending long hours in the library and making notes thereof. When modern technology is available for comfort, it would be unfair to say that the students should not avail thereof and continue to study as in ancient era. No law can be interpreted so as to result in any regression of the evolvement of the human being for the better. (p. 84).
Social advocates hailed the verdict, saying the court had correctly upheld the supremacy of social good over private property. Students had rallied behind the photocopier, saying most of the books were too expensive.

The publishers plan to appeal, arguing that the trial court’s approach goes far beyond any reasonable interpretation of the exception in the copyright act for educational copying.

Stay tuned for next year’s updates of these fast-changing legal areas.
Is a Gold Open Access World Viable for Research Universities?

Greg Tananbaum, ScholarNext

Carol Tenopir, University of Tennessee and Hanken School of Economics, Helsinki, Finland

Ivy Anderson, California Digital Library

Abstract

Open access is at the heart of a seismic shift in scholarly publishing. In particular, gold open access (OA) has expanded at an accelerated pace, increasing in market share every year. In the gold OA model, financial viability shifts from the demand to the supply side, with article processing charges (APCs) a common scenario. Ideally, this model would be sustainable for academic research institutions, in that it would cost them cumulatively no more to pay APCs than they pay now in the traditional subscription model. APC-driven gold OA has financial and other implications for libraries, institutions, and authors. In the Andrew W. Mellon Foundation-funded Pay It Forward project, we examined the viability of gold OA by looking at institutional costs, faculty and graduate student opinions, and various models for gold OA. The Pay It Forward research teams gathered a variety of qualitative and quantitative data from publishers, research libraries, and faculty and students including current APC charges, current subscription charges, journal publication costs, opinions and behavior of graduate students and faculty members regarding publishing, reading, and OA.

Project Impetus

The Pay It Forward project began with an observation. It has become increasingly clear over a period of years that North America is primarily moving in a green open access (OA) policy direction, while policy developments in Europe and the U.K. are driving a conversion to gold OA. In other words, two of the largest research publishing economies in the world are working potentially at cross-purposes when it comes to open access developments. Together the United States and Canada comprise 31% of worldwide output, while Europe and the U.K. comprise 34%, meaning these developments could in fact be on a collision course.

These trends appear to be setting up a confused economic situation. License fees and article processing charge (APC) revenues are increasing; double-dipping opportunities abound through hybrid journals. Gold OA is now about 15% of all publishing worldwide, and it is projected to continue to accelerate rapidly over the next five years (Björk et al., 2010; Laakso & Björk, 2012; Laakso et al., 2011). Therefore, the project principals agreed that it would be useful to have a firmer grasp of these trends’ implications.

The California Digital Library (CDL) began this process in an informal manner. In 2013, at the request of the university librarians, CDL created some preliminary modeling of what the impact of a conversion to gold OA might resemble. After all, CDL licenses most of the journal content available system-wide at the University of California (UC). It had also been purchasing customized reports from Thomson Reuters about UC publishing rates in the journal packages that they license to inform their journal negotiations, so we had a good base of both financial and authorship data from which to work.

Preliminary calculations from this exercise were intriguing. They suggested that far from saving money if the world suddenly flipped to gold OA, the University of California might, in fact, spend more money on scholarly publishing but that gold OA might be affordable under certain conditions. Therefore, we began to wonder (A) if this quick and dirty analysis was correct, and (B) if it might also be true for other large, research-intensive institutions.

One of the challenges in exploring these issues is that there is plenty of opining about the viability of open access but much less objective analysis. Therefore, dispassion had to be a key pillar of the project. We wanted to stay away from questions such as, “Would society be better off in a fully OA world?” instead focusing on very practical, data-driven considerations.
Project Goal

The goal of this project can be distilled to one primary question: Can such a shift to gold OA be viable and financially sustainable for the institutions that publish the lion’s share of research in the United States and Canada? It is important to note that we defined sustainability in this project as “costing those institutions roughly no more than, and ideally considerably less than, current journal subscription costs for comparable journals today, with a rate of growth that will be possible for these institutions to support over time.” No matter how attractive the economics of OA might look from the perspective of the scholarly system as a whole, no institution will be incentivized to move in that direction if it isn’t sustainable on a local level. As one of the largest public research institutions in the world, with a significant publishing profile—UC publishes something like 2% of the world’s research literature—we had a real curiosity to figure out if a fully gold OA environment could be viable from the perspective of the big research school.

Team and Partner Roles

Because we did not want the project to be UC-centric, we engaged a set of partners from public and private institutions who share UC characteristics of high publication output. Thus, the Pay It Forward project includes Harvard University, The Ohio State University, and the University of British Columbia. The core project team consists of MacKenzie Smith, UC Davis, University Librarian and Project PI; Laine Farley, CDL Executive Director and Project co-PI until her retirement; Ivy Anderson, CDL Director, Collection Program and Project Quantitative Lead; Mathew Willmott, CDL Data Analyst; Carol Tenopir, University of Tennessee, who conducted the author opinion and behavior studies; David Solomon, Michigan State University and Bo-Christer Bjork, Hanken School of Economics, responsible for APC research; economist Mark McCabe, Boston University, who ran point on scenario modeling and economic analysis; and Greg Tananbaum, who served as project manager and contributed to the publishing economics section.

We had the further support of two industry partners: Elsevier and Thomson Reuters. They helped us directly with both bibliometric analysis and the provision of raw data, broken out by discipline, about both worldwide and institution-specific publishing outputs. Finally, the Association of Learned and Professional Society Publishers (ALPSP), a society with some 300-plus member organizations including both the large commercial publishers and society publishers, assisted us with the publisher survey to gain a better understanding of publisher attitudes and strategic directions with respect to OA.

Key Project Components

The project was built upon qualitative and quantitative components, each data driven. We took this approach to ensure that we were not driven purely by economics but also took into account social and behavioral dynamics and values. From the qualitative perspective, we performed extensive survey and focus group work with faculty, grad students, and post-doctorates. We also worked with the Association of Learned and Professional Society Publishers to survey its membership. This gave us interesting and useful information about publisher attitudes and activities related to open access.

From the quantitative perspective, we performed a much richer and more detailed elaboration of the kind of modeling we had done earlier, examining publishing output and licensing costs under a variety of scenarios that were then informed by detailed research and analysis. Among the areas we delved into was a five-year deep dive into what the partner universities spent on scholarly journals from 2009–2013. We also thoroughly examined our partner universities’ faculty publishing activities, including co-authorship patterns, availability of research funding, and growth over time, for this same five-year window. Additionally, we explored what the true cost of publishing is under the current environment by looking at dozens of publisher tax documents, real-world APCs for fully OA publishers, and previously published literature and analysis of this issue. Taken together, these data helped us build a set of financial scenarios, or models, depicting the implications an APC-based system of scholarly journal publishing for large research institutions. The Pay It Forward final project report may be found here: http://icis.ucdavis.edu/wp-content/uploads/2016/07/UC-Pay-It-Forward-Final-Report.rev_.7.18.16.pdf

What Do Faculty and Students Think of Gold OA?

The Author behavior team’s role in Pay It Forward was to measure attitudes toward and knowledge of
gold open access (OA) among faculty and graduate students at participating research universities. To capture this information, we held focus groups in 2015 at The Ohio State University, Harvard University, University of California Davis, University of California Irvine, and the University of British Columbia. Each location held two focus groups, one for faculty and one for graduate students, and there was a total of 77 participants with 46 faculty members and 31 graduate students. These focus groups helped us to devise a survey that was then distributed at four of the five institutions. The survey had 2,021 responses for a response rate of 14.1%. The survey respondents were almost evenly split between faculty members (46.3%) and graduate students (45.3%), with a few post-doctoral researchers (8.4%). Of the graduate students, 80% were PhD students. Respondents were generally evenly distributed among subject disciplines, with slightly more coming from STEM disciplines. We also had a wide range of career ages. For faculty, the average year that they obtained their highest degree was 1955 (with a range of 1959–2015), graduate students was 2016 (with a range of 2012–2023), and postdoctoral researchers was 2012 (with a range of 2002–2015). Almost all respondents had published articles in the last three years.

Attitudes Toward Gold OA

There is a wide range of opinions about gold OA, from the quite positive to the quite negative. This observation became evident early in the focus groups and then was clarified in the survey responses. The following comment is typical of those holding positive views of gold OA: A graduate student stated, “It matters heavily to me that my papers are open access. From my value standpoint, I care less about the impact factor, and I care more about having it peer reviewed but open access.” A few faculty members said that they would only publish in OA; the most common reason behind this decision was that it is more ethical to make result of research open. Several stated that they wanted to make sure that those without access to large library collections could still access research.

Table 1. Respondents’ subject disciplines by position type.*

<table>
<thead>
<tr>
<th>Subject Discipline</th>
<th>Position</th>
<th>Overall Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Faculty</td>
<td>Graduate Student</td>
</tr>
<tr>
<td>Arts &amp; Humanities</td>
<td>197</td>
<td>149</td>
</tr>
<tr>
<td></td>
<td>56.3%</td>
<td>42.6%</td>
</tr>
<tr>
<td>Engineering &amp; Computer Science</td>
<td>80</td>
<td>170</td>
</tr>
<tr>
<td></td>
<td>29.9%</td>
<td>63.4%</td>
</tr>
<tr>
<td>Life Sciences &amp; Medicine</td>
<td>315</td>
<td>208</td>
</tr>
<tr>
<td></td>
<td>50.6%</td>
<td>33.4%</td>
</tr>
<tr>
<td>Mathematics</td>
<td>28</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>62.2%</td>
<td>26.7%</td>
</tr>
<tr>
<td>Physical Sciences</td>
<td>77</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>44.0%</td>
<td>42.9%</td>
</tr>
<tr>
<td>Social Sciences (including Business, Education, &amp; Law)</td>
<td>236</td>
<td>293</td>
</tr>
<tr>
<td></td>
<td>43.0%</td>
<td>53.4%</td>
</tr>
<tr>
<td>Other</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>10.0%</td>
<td>80.0%</td>
</tr>
<tr>
<td>Total</td>
<td>934</td>
<td>915</td>
</tr>
<tr>
<td></td>
<td>46.3%</td>
<td>45.3%</td>
</tr>
</tbody>
</table>

*Only two respondents did not answer this question regarding position type or subject discipline.
On the other end of the spectrum, some respondents conflated gold OA with vanity publishing. One respondent explained, “If a particular venue becomes associated with a vanity press—if you have enough money, you can get it published there—then it loses credibility in academic circles or elsewhere.” Many of the concerns that did not conflate OA with predatory publishers largely saw article processing charges as a barrier to publishing, for themselves and for others. Most opinions, however, were neutral, and the more neutral attitudes toward gold OA may be better characterized as apathy. They had not thought much about the cost of publishing or prices of publications but instead, focused on publishing their research in the highest quality venue possible.

Factors in Determining Publication Outlets

In the survey, we asked respondents to rate the importance of a variety of factors in choosing a journal to which to submit or publish their work. Respondents were asked to rate the importance of each factor on a scale of 1–5 (1 = not important; 5 = very important). They were also given the option of “not applicable.” Open access rated the lowest in importance across all position types and subject disciplines (Table 2). Our recent article in Publications examines author motivations in choosing publication outlets (Tenopir, Dalton, Fish, Christian, Jones, & Fish, 2016).

Table 2. Ranking the importance of journal factors.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality and reputation of journal</td>
<td>4.69</td>
</tr>
<tr>
<td>Fit with scope of journal</td>
<td>4.61</td>
</tr>
<tr>
<td>Audience</td>
<td>4.49</td>
</tr>
<tr>
<td>Impact Factor</td>
<td>4.09</td>
</tr>
<tr>
<td>Likelihood of acceptance</td>
<td>3.74</td>
</tr>
<tr>
<td>Time from submission to publication</td>
<td>3.58</td>
</tr>
<tr>
<td>Editor or editorial board</td>
<td>3.42</td>
</tr>
<tr>
<td>Open access</td>
<td>2.84</td>
</tr>
</tbody>
</table>

* N = 2021

Perhaps because of the perceived stigma of “pay to publish” or predatory journals, or perhaps because the issue of open- or subscription-based journals did not resonate with many respondents, for most OA was not an important factor when choosing where to publish (Table 3.)

Although half of the respondents agreed or strongly agreed that more people would read and use their research if it were published in an OA journal (50.2%), almost as many felt that article processing charges (APCs) would limit their ability to publish (46.2%). Consequently, 40% of respondents would find other ways to publish. Only 33% of respondents agree that APCs are a reasonable alternative for publishing in an OA journal. Very few (14%) of respondents believe that APCs reflect the quality of a journal.

Table 3. Percentage of respondents’ agreement.

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>More people would read and use my research.</td>
<td>764</td>
</tr>
<tr>
<td>APCs would limit my ability to publish.</td>
<td>718</td>
</tr>
<tr>
<td>I would find alternative ways to publish.</td>
<td>571</td>
</tr>
<tr>
<td>APCs are a reasonable alternative.</td>
<td>504</td>
</tr>
<tr>
<td>APCs reflect the quality of the journal.</td>
<td>185</td>
</tr>
</tbody>
</table>

How Much Are Researchers Willing to Pay?

We asked respondents how much they would be willing to pay in APCs based on different sources of funding, such as personal funds, discretionary research funds, OA publication fund through the library, department or other institutional research funds, grant funds, and other nonspecified funds. The majority indicated that they would be willing to pay somewhere between $0 to $499 (Figure 1).

Paying from personal funds is clearly unpopular. The library was the only source indicated by more than one-quarter of respondents to pay between $2,000 to $2,999. This fee amount is more typical in the sciences; therefore, this chart needs to be put into perspective. Half of the journal article publications in these universities came from the life sciences and medicine. Those scientists (32.2%) are more willing and more accustomed to paying $1,000 or more from their grant funds than researchers in other disciplines, yet only 19.4% of physical scientists, 12.6% of engineers/computer scientists, 9.9% of social scientists, 9.7% of mathematicians, and 4.6% of arts/humanities are willing to pay $1,000 or more.
from grant funds. Our College & Research Libraries articles examines more closely the demographic differences in author willingness to pay APCs by funding choice (Tenopir, Dalton, Christian, Jones, McCabe, & Smith, 2016).

**Demographic Differences**

There are other demographic differences as well. Applied STEM fields such as engineering and medicine are more accepting of OA, but they also care more about impact factor. On a scale of 1 = disagree strongly and 5 = agree strongly, engineering/computer scientists ($M = 4.17$) and life sciences/medicine ($M = 4.15$) rate impact factor higher than the social sciences ($M = 4.08$), physical sciences ($M = 2.68$), humanities ($M = 3.95$), and mathematics ($M = 3.50$). Scholars in the humanities (their own ability to publish = 3.59; others’ ability to publish = 4.25) and social sciences (their own ability to publish = 3.40; others’ ability to publish = 4.17), on the whole, worry more that gold OA fees will hinder their ability as well as others’ abilities to publish. The level of agreement from respondents in the mathematics, physical sciences, engineering/computer sciences, and life sciences/medicine ranged between $M = 2.99$ to $3.12$ for their own publishing opportunities and $M = 2.97$ to $3.17$ for potentially limiting others’ publishing abilities.

Graduate students ($M = 2.99$) and post-doctoral researchers ($M = 3.29$) are slightly more likely to agree or strongly agree than faculty ($M = 2.80$) with the statement that APCs are a reasonable alternative to subscription fees. On the other hand, compared to graduate students and post-doctoral researchers, faculty are less likely to think that OA will increase readership or the quality of research. They are also more likely to equate OA with lower quality research.

The quality and reputation of a journal is still what matters the most to academic authors and quality is most often defined by traditional measures. These qualitative results together with the wide range of quantitative data collected helped the teams shape potential solutions.

**Article Publishing Costs**

In our cost-per-article analysis, we attempted to ascertain what a sustainable journal publishing operation might cost on a per-article basis. We first explored the possibility of constructing a ground-up cost model. This was ultimately dismissed as unfeasible for a variety of reasons, notably the high degree of variability in what constitutes publishing services. In its place, we examined actual cost data from a variety of sources, including tax forms, literature reviews, analysis of gold OA journals in which our authors publish, and discussions with publishers. This process allowed us to develop a floor and average cost per article, including a 13% surplus to fund ongoing innovation. This sustainability range, from $1,103 at the low end to $2,566 at the high end, helped to establish the viability of the financial model we developed and test whether it could provide sufficient income for publishers to sustain their core functions.

![Figure 1. Willingness to pay APCs by funding source.](image-url)
Complementing our cost-per-article analysis, various types of APC data were gathered for a thorough analysis of publisher and author behavior in setting and paying APCs. List price APC data for full OA journals were gathered from a longitudinal study led by Heather Morrison and were updated by our own investigations. We mapped the pricing dataset to our publication output data set to estimate how much researchers at our partner institutions paid in APC charges for publications in existing full OA journals over the course of the study (~$1,892), as well as the average APC set by publishers for journals in which authors at our partner institutions published (~$1,864). Additional data gathered from various European databases recording actual APC payments made by granting agencies or institutions on behalf of authors corresponded well to the average APCs determined in our partner mapping exercise (average $1,865 for publication in a full OA journal).

**Modeling Future APCs**

Analyzing current APCs was instructive about the APC market as it exists today but was not sufficient to help us understand how APCs might evolve in the future if such practices were to become the norm. Given the findings from our author research about the importance of journal quality (as perceived by the author) in publication decisions, we approached this question through an economic analysis of the relationship between price and journal quality, using the journal source normalized impact per paper (SNIP) values as a proxy for journal quality. Our hypothesis was that in a true APC market, competition for authors will lead publishers to price their APCs based on a journal’s perceived value to authors, which in turn will turn on perceptions of quality. A linear regression performed on a subset of APC pricing data, narrowed to journals from publishers that employed differential APCs for their journals, revealed a correlation coefficient of 0.654 based on SNIP quality values. The equation generated by this regression allowed us to predict the APC of any journal, given that journal’s SNIP value. The APC for a baseline journal in this analysis (SNIP = 1.0) turned out to be $1,857, in line with the average APCs uncovered elsewhere in our study. We then used this equation to predict the APC for every article in our bibliometric data set, thereby calculating the total cost of each institution’s scholarly publishing activities for each year in our study.

**Affordability of an APC Model**

Our project defined affordability in terms of the relationship to current licensing costs: Would an APC-driven model be more or less expensive than a library’s current journal subscriptions? We examined this question by calculating an APC “break-even” point for our library partners—what level of APC could each partner afford given its publishing output—and comparing that with the averages identified in our study. As one would expect, affordability turned on the research productivity of each partner. Smaller, less research-intensive institutions with lower publication output would be likely to realize substantial savings under an APC model, whereas the larger institutions would be likely to see their costs increase. For all our partners, given their research characteristics, an APC model would exceed the capacity of their current library budgets, significantly in some cases. However, the availability of grant funding changes this picture dramatically. Grant-funded research was another parameter analyzed in our study. By identifying all articles that included a grant acknowledgement statement, an attribute that is tracked in the Web of Science bibliometric dataset, we were able to estimate the number of articles for which sponsored research funding might be available to cover an APC. In fact, we know that most articles being funded via APCs at our institutions today are paid for in this manner. A large percentage of our partner institutions’ sponsored research funding (~72%) comes from federal agencies whose policies treat publication costs as an allowable expense, and many private funders have adopted such policies as well. When articles acknowledging a grant were eliminated from the total, subventing APCs for the remaining articles proved to be within the current library budget for even our most research-intensive partners.

**Can APCs Be Made Sustainable?**

Even if APCs are envisioned to be affordable under certain conditions today, a key concern in modeling a potential APC future is how to control costs and make them sustainable over time. Libraries’ experience with runaway journals inflation is a cautionary lesson that would be important to guard against in designing a financial model for APCs. We developed a set of five criteria for a financial model...
based on economic theory and the conclusions drawn from our author focus groups and survey: Libraries should continue to play a major funding role in any scenario; grant funding should be considered a legitimate and routine source of funding for open access publication charges; establishing the right marketplace incentives should be a key component of any funding model; to achieve a functional incentive structure, authors should have “some skin in the game”; and authors should not bear an undue burden in an APC-driven model. A fundamental premise is that a properly functioning journals marketplace requires author participation, rather than the purely intermediary relationship that obtains between libraries and publishers today.

**Multipayer Approach**

The result of our modeling was a multipayer strategy in which libraries and their parent institutions, authors, and funding agencies all play a role. Libraries would establish a baseline of APC support to ensure that authors are not overly burdened, monitoring the marketplace to determine appropriate levels of funding. Authors would be required to “top up” this subsidy when necessary, utilizing either grant funding or other institutional funds that would be made available to them for publication support (and potentially other purposes). Authors would be naturally incentivized to economize in their use of these funds to stretch their research dollars, thereby exerting pressure on publisher pricing that would restrain or even lower APCs over time. Our modeling of this scenario suggests that distributing costs in this way would indeed be cost-effective for the large research institutions in our study.

**Conclusions**

As open access business models continue to evolve, libraries must plan for the significant impact of these changes on their budgets and professional practices, and they must seek to shape the new world that is emerging. While we do not yet know how fully open access publishing will take hold or what business models will prevail, the APC model has emerged as a leading contender for much of the Western canon and warrants our close scrutiny. In North America, library journal budgets alone will not fully cover APCs for research-intensive institutions in a flipped gold open access world. However, grant funding received by authors at those institutions would cover this difference in the vast majority of cases. In addition, our research suggests that involving authors in payment decisions by making discretionary publication funds available to them would introduce APC price competition, without interfering with author choice in where to publish. This would encourage a competitive journal market and drive costs down over time.
These conclusions and the modeling done in our Pay It Forward project, while both rigorous and intriguing, remain a set of hypotheses to be tested in the cauldron of experience. We are continuing to explore these issues and plan to seek out opportunities to test our ideas as the scholarly communications environment continues to unfold.

References


