

Update on Industry Trends and Issues

Gary Price
InfoDOCKET

Rick Anderson
University of Utah

Maria Bonn
University of Illinois

Follow this and additional works at: <https://docs.lib.purdue.edu/charleston>

An indexed, print copy of the Proceedings is also available for purchase at:

<http://www.thepress.purdue.edu/series/charleston>.

You may also be interested in the new series, Charleston Insights in Library, Archival, and Information Sciences. Find out more at: <http://www.thepress.purdue.edu/series/charleston-insights-library-archival-and-information-sciences>.

Gary Price, Rick Anderson, and Maria Bonn, "Update on Industry Trends and Issues" (2016). *Proceedings of the Charleston Library Conference*.
<http://dx.doi.org/10.5703/1288284316503>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

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 to 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 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

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 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, I used to work in the University Press world. We talked a lot about the "educated general reader" where it would've been more simple to say probably, "Yeah, right," but maybe there is an "educated general librarian" who follows this world and cares about it but may not pay attention to the local details like Gary does. And so looking at something like this, I'm struck with wonder and bemusement of what can we do with

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 here's a presentation, a panel that he was on in June. 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 [pop-up archive](http://pop-up.archive.org) 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](http://pop-up.archive.org) 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, Time Traveler 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.

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, and 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 try 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.

Gary 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.