The Long Arm of the Law: Privacy Explored

Gary Price
infoDOCKET

William Hannay
Schiff Hardin

Lisa Macklin
Emory University Library

Ann Okerson
Center for Research Libraries

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, William Hannay, Lisa Macklin, and Ann Okerson, "The Long Arm of the Law: Privacy Explored" (2015). *Proceedings of the Charleston Library Conference*.
<http://dx.doi.org/10.5703/1288284316229>

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.

The Long Arm of the Law: Privacy Explored

Gary Price, Cofounder and Editor, infoDOCKET

William Hannay, Partner, Schiff Hardin

Lisa Macklin, Director, Scholarly Communications Office, Emory University Library

Moderated by Ann Okerson, Senior Advisor on Electronic Strategies, Center for Research Libraries (CRL)

The following is a transcript of a live presentation at the 2015 Charleston Library Conference.

Ann Okerson: Well, good morning and welcome to the sixth session of the “Long Arm of the Law.” Thanks to Katina for thinking these topics are so important that she makes sure that we get a certain amount of airtime for them each year. Every year we pick a topic or a court case or something that we’re going to follow in this session that will be of value to many, many in the audience: librarians, publishers, and others. This year we decided to focus on privacy, which is a topic that is touching us all in every layer of our lives. Privacy is the ability to control into whose hands information falls. It is important because it protects the rights and freedoms of all of us to live as we please. Our speakers today are bringing to us important perspectives about privacy in the library, in publishing, and other settings. I’m going to be very brief with these introductions. We’re going to just run through them all very quickly. The speakers will each take about 15 minutes, and then with good luck we’ll have some time left over for audience discussion.

Our first speaker is Gary Price, who is known to you in many ways for his work with *Library Journal*, keeping us all informed and up to date with infoDOCKET. He is going to speak about technologies that libraries and publishers use—deploy, often inadvertently—and how they might compromise our privacy. The other day Gary told me that my most frequently used airline, United, doesn’t encrypt their itinerary information enough and it might be possible for people who care to do this to find out where I’m going, and that freaked me out. So, Gary is going to talk about things like that. His mission here is to put the fear of God into our minds as we go forward. Bill Hannay, who

is a repeat, regular offender here and one of our most popular speakers, is going to talk about the larger legal context, continuing a topic that he started last year, which was the right to be forgotten. And then Lisa Macklin is going to bring the discussion directly into the academic setting and discuss things that will be of issue in our library context. So, that is our lineup for this morning, and we thank you for being here. We hope there is something in here of value to everyone. So, Gary . . .

Gary Price: All right. Thank you, Ann. Good morning, everyone. To continue with the Kenny Rogers theme, hopefully by the end of this you won’t call me “Coward of the County.” Okay. I’m here to put the fear of God in you, and hopefully in the next 12 minutes we’ll do some of that and also impart some useful knowledge to you as well. So, library and Internet privacy. Bottom lines at the top: awareness by all parties, discussion, and education is a must and needed now. The privacy that people expect from the library, and have come to expect from the library over many, many decades, is not where it should be in the digital age and we need to do more. The first place to begin is with discussion, education, and knowledge of first ourselves and then our users. Some of the good news is that there is work being done by many organizations. Now in the last couple of years we’ve seen interest from NISO. They have a privacy initiative. The American Library Association just released e-book privacy guidelines (<http://www.ala.org/advocacy/library-privacy-guidelines-e-book-lending-and-digital-content-vendors>), for example, and there is another funded project from the Knight Foundation called the Library Freedom Project. We are also seeing more publishers getting involved. Some examples, and database

providers—BiblioCommons, Project Muse, OverDrive, BilblioBoard—are all now encrypting the entire process, which includes the personally identifiable data (your name, if you have to register), but what I'm talking about is also the transmission of data over the Internet, and we'll talk more about that in a moment.

Privacy is more than encryption. It is the local data: How is it stored? The transaction laws: Is the data being scrubbed? How often? Does your library—and I've seen this in some of my research, privacy policies from some libraries haven't been updated in years: Is your library doing that? Are you disclosing what's going on in a clear manner, not on page 43 of some statement? What are we doing with sharing with vendors? We can share with vendors some of the things that Katherine [Skinner] was talking about—we can use some of this data—but we need to let our users know what's going on. We ask transparency of others, but we're not doing the job that we're asking of others. We need to do a better job. The same thing is going with data storage. There are third parties involved. The library is just not a single building now. The data is flying all over the place. What is your relationship with the vendors? What are their privacy policies? How will they contact you? Will they contact you? How long after might there be a data breach? Is the technology configured correctly? Opt-in services the vendors provide: What are they doing with that data? If you're e-mailing results to a group of people, is that data being saved? If it is, is it being scrubbed after x amount of days? Is it just immediately disappearing? We need to understand these things and work with our vendors on these types of issues. Is there a response plan in place locally that, if there is a data breach, what will happen? So, it's awareness, training, staying current, and being vigilant because all of this changes on a very regular basis.

I also have seen over the last few years that perhaps we're a bit scared of what might happen, but in my view that's kind of burying your head in the sand. And in this day and age we can lose a lot of respect from the general public and our users for something that they respect us for—being a pillar of privacy in our communities—with just one

simple data breach, and the fact that somebody can easily say, “Well, you never told us.” I believe in this day and age we need to be out front on these issues and, again, the transparency that we ask of others, we need to become better at in this particular topic. We need to be preemptive versus reactive. If just about every European website can have a cookie disclaimer on it, something that simple, why aren't we doing the same thing? For example, when somebody borrows a book from an e-book provider and puts it on their Kindle, that data is being shared with a third party. Are we simply even disclosing this to our users? And one thing that I've also learned over the last few years is everybody has a different level of comfort with privacy. So, if somebody doesn't want their data stored indefinitely with Amazon, are we showing them how to remove the data off of the Amazon server? Are we providing them that information? From what I've seen we need to do a better job.

Now, we can talk about all the ethical reasons and “this is what we have always done” reasons, but put all of those aside for a moment. In the last few years, privacy, data surveillance, and the like have become a major topic both in the US and around the world. My belief is that privacy, library privacy and privacy in general, can be a wonderful marketing opportunity for us as a community to show our relevancy—something we are always looking to do—with an issue that a lot of people are talking about and interested in. For example, the community campus can be a privacy clearinghouse and leverage respect for our past efforts, and public respect.

I'm going to skip through these [slides] and I'm going to show you a little bit of a live demo, so let me scroll through this here, and we'll go over here. So here it is, ladies and gentlemen. Your term for this part of the presentation is *sniffing*, and this is an open-source-type piece of technology that I am, on a level from 1 to 10, on a scale from 1 to 10, I am probably a 2 or 3 at. You can take intensive classes on this. This tool called Wireshark has conferences that keep people updated. In a nutshell, this is all of the Wi-Fi traffic flowing through this room right now. You can break it down and actually see sometimes, if it is

unencrypted, what somebody is doing. So I will now, with the next tool I'll show you, it makes it a little bit easier to see what I'm talking about. This is a piece of software—again, open-sourced—developed by a student at Iowa State University. Instead of writing a master's thesis paper he developed this software. This takes all of the data coming off of the unencrypted cookies coming through Wireshark and makes them much more easy to view without having to have a master's degree in that particular product. And this is how—one of the tools—that I learned that Ann's airline and another European airline is sending data unencrypted—including the passenger's last name and their passenger record number over the Internet unencrypted—which I found to be rather shocking.

Here are all of the people on Wi-Fi on the network right now. There's probably more, but this is what we're seeing right now. Now, you're going to say, "Well, Gary, so what if I'm learning that you're searching x library and it says that you're looking up whatever?" Well, another thing that's going on right now, especially when you're thinking of government surveillance and surveillance by others, is that whatever you're doing—somebody is searching books.Google.com—this tool also is identified with the exact what is called the MAC address, which is the unique ID of that person's computer or electronic device, so there's two pieces of data. There is no one holy grail of data. In this day and age, having enough distinct data points makes it relatively simple to figure out who that specific person is. So, you can see now what people are looking at—somebody's looking at the CBC—that kind of thing. So those are the two particular tools.

Now, let me go back to my slides for a moment and give you a couple of actual examples. Am I scaring the heck out of you? This is one slide, if we can do this here. Can you see this right here? I'm not going to make this one too big. Here is a practical tip for you—especially for those of you who own Apple devices, very often your device is named after you. So, one of the things floating through is not just that MAC address but it's your specific name. You can see that here. (Play from current slide.) You can see that somebody named

Jeff W., I blanked out his full name, was on the Wi-Fi yesterday at one of the conference hotels. Do you see where I'm going with this? It's these distinct, a number of these distinct data points merged together, which can be easily done now, makes it very simple to figure out who someone is and where they are.

Another thing that we're dealing with now all over the Internet are things called tracking scripts and beacons. These are pieces of code embedded in the webpage that can provide all sorts of data. Again, these are the types of analytics we were talking about earlier—whether or not it is a good thing or bad thing, that is up to the organization—but my point is, gaining that knowledge and sharing what's going on with our users. Also now we are seeing beacons used in a mobile setting, allowing people to be tracked throughout the store or throughout the library. If your library is doing that, what are you doing with that data after the person walks out of the library? And are you informing them in the first place that that's going on?

Your IP address can share a lot of information with you. Now in this, case at my hotel, you're able to see specifically where I am. This IP address is tied to Charleston, South Carolina, but I've also been in locations where it not only says that you're in this city but where in this city you're at. You're in the lobby, for example, of the Mayflower Hotel in Washington, DC. This is again more of a specific example of the Cookie Cadger tool I just showed you. Here I am, 9cf3, and here I am searching the library of Congress OPAC, and you can see all of the search suggestions as I'm typing in. All of that is being sent over the Internet unencrypted. Here is another view of that data. You can see the referrer says I was looking at loc.gov, and then my search term. This is another example, and in this case the Google Analytics we're sharing show that this person was coming from library.caltech.edu. This by itself isn't a bad thing, but again my point, my takeaway from this, is we need to understand what's going on first and then share that data with our users.

Here is an example of a PubMed search. You probably can't see it from here, but you can see all

of my search terms and the specific data fields I was searching. So, become privacy literate, first step. It needs to be a major part, in my view, of any digital literacy campaign that you're doing. Awareness of tools and concepts to minimize exposure: So I've shared some of those with you. I know Lisa [Macklin] is going to—hopefully we will have time to talk about some of the things that you can do if this is an issue for you at the library level. But also for an individual, remember everybody has a different threshold of comfort. There are tools that you can use, but while you can't—I would never get up here and say that you can be completely private online and I can guarantee it—but there are tools to, let's say, put “roadblocks” up to make it a little bit more challenging to get to some of this data. It's not rocket science to use them.

Stay current: Things change quickly, we all know that. And discuss issues with colleagues and our users. There are a lot of ethical issues, which we will get into moving forward and then teach our colleagues and users. I've also on the slide deck that you'll get included some other examples. This for example: the article *Exposing the Hidden Web: An Analysis of Third-Party HTTP Requests on 1 Million Websites* (<http://arxiv.org/abs/1511.00619v1>) was just published in the past month. So, again, I promised I would go a little bit short to stay on schedule. I want to make Ann a happy lady. One final comment: These types of tools, Wireshark and Cookie Cadger, don't just work on Wi-Fi. You could plug an Ethernet cable on here and get the same type of data as well. So, we will have questions later. Thank you very much for your time and attention. I hope I've scared you just a little bit. Thank you.

Bill Hannay: The title of my program is “Please Remember to Forget.” Catchy, right? I'm sure you will remember because this program is all about forgetting. And the larger topic, where's all this privacy stuff going? Is it going to turn into a flood, or is it just a ripple? We'll talk a little bit about this.

As I mentioned last year, Europe, the European Union, is very big into privacy issues. In particular, unlike the United States, they have a very broad-reaching data privacy statute. It's called a

“directive,” but it's effectively a statute. And its idea is to protect individuals in the processing of data about them. What they focus on is personal data. Now, personal data actually has a very broad definition in Europe: Any information relating to any individual. Any individual! Any information! Whether it relates to personal, private, or professional life. It can be a name; it can be a photo; it can be an e-mail address. They don't have Social Security numbers over there, but if they did, Social Security numbers, medical history, really anything about them, particularly if it happens to be of a deleterious nature. And why are they concerned about this? A great quotation from the memo written in the European Union government, “With social networking sites, cloud computing, location-based services and smart cards, we leave digital traces with every move we make.” Gary just showed you some of your traces: all you in the audience just buzzing around the Internet when you're supposed to be listening to us. “We need a robust set of rules to make sure people's right to personal data protection . . . is made effective” (EU Memo/12/41, Brussels, 2012, http://europa.eu/rapid/press-release_MEMO-12-41_en.htm?locale=en, para. 1).

I kind of admire the European Union government because they have apparently a sense of humor. This is actually an official government cartoon that's on their data privacy website. It looks like a student going into the free Internet café and coming out embarrassed and totally naked because of the information that this person has logged onto the Internet doing an Internet search. I am reminded of going into a library one time a couple of years ago and waiting in line to speak to a human being who was a reference librarian. And there was a young person, who looked like a high school student, standing, talking, kind of whispering to the reference librarian. Whisper, whisper. And the librarian leans back and says, “Mildred! Where do we keep the books on sex change operations?!?” So, information about oneself can be disclosed in a number of ways.

You may remember last year we talked about a man named Mario Costeja Gonzáles. And he asked Google to take down some information about him—that is, remove it, as in take down,

because this deleterious information, this negative information about him had long since ceased to be accurate and it was embarrassing. If you did a search on his name, it would come up with a couple of articles about the fact that he had, well actually I won't mention it because I don't want to publicize what his problem was. He wanted this taken down, and they said, "Oh no, we couldn't do that." So he goes to the government agency in Spain and asks them to make Google take this down. They enter an order, Google appeals, it goes all the way to the European Court of Justice, which is, on a regional basis, like our United States Supreme Court; it's the highest court in Europe. The Court affirms the order of this Spanish privacy agency and makes Google take this down. But more than that, not just fix the problem for Sr. Costeja Gonzáles, but to actually order Google to set up a whole system to do this for anyone who is subject to the data privacy laws, or the protections of it. What they ordered was that Google remove from the list of search results any webpage links relating to an individual if such information is "irrelevant" in relation to the purposes for which data is collected. You have to balance individual rights against the collective rights of the public to know things.

Google has set up this system, and there's a request form here that you can go on the website and say, "I'd like you to remove the following information." And they have been, this process has been used an enormous amount in the past year. As of this week, 338,000 requests have been filed with Google in Europe to remove information from 1.2 million URLs. Of those, 42% have actually been removed, these search results. Now understand, what that means is, it's not whatever the URL is has been removed from the Internet, it's that Google's results will not produce that URL. You can maybe still find it some other way if you happen to know it, but you can't search for it. So, almost 10,000 URLs, the listing of them was removed from Facebook.

For example here, here's one that in the UK someone said that I'd like you to remove a minor crime that I was involved with. Google did remove it, and then the newspaper that had reported it

originally then filed a news report about the removal of the report, and Google took that one down too. So the long arm of the law extends on a continuing basis.

The EU is actually thinking of strengthening this whole process and reaffirming the right to be forgotten. They prepared an elaborative directive in 2012, just a few months ago the Council of Justice Ministers approved it, and it will now go to the EU Parliament and the broader Council for further action, and I expect that it will be enacted. It addresses a number of things, not just right to be forgotten, but they have reinforced the notion that if an individual no longer wants their personal data processed, and there's no legitimate reason to keep it, then it should be removed. But, this proposal recognizes something that was somewhat unclear under the prior codification, which is that this is all subject to a careful balancing of the right of freedom of expression, for newspapers to print articles. Even though a newspaper may be required to take it down, or Google's search will not find it, that has to be taken into account.

Here's a statement recently from the EU Justice Commission, sort of the equivalent of our attorney general:

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. (European Commission, 2015)

Again, a balancing of the public's rights and private rights.

To update you on this thing: Not only are there the hundreds and thousands of requests that are going in Europe, it now is beginning to impact in the United States because a French agency has ordered Google to take down information—to remove, to "delist" information—from its US-based website, from Google.com; not just Google.es or Google.fr in Europe, because they've been doing that. But now this French court says in

order for this to be fully effective, this whole right to be forgotten thing, you have to take it down everywhere that you, Google, have a website. And so if you've got Google.com in the US, it's got to come out of that, too, because someone in Europe could link to Google.com and do the search, which would otherwise be prohibited in Europe. The logic of this is, it's an evasion otherwise. Well, Google is not taking this lying down. They're going to appeal to the European Court of Justice again. They've said, "We're not going to do this. We're going to take the heat and we're going to go and appeal because you don't have any jurisdiction." And what the French agency says is that it isn't a question of jurisdiction over the United States—what we're saying is if you've got a website somewhere else and someone in Europe can link to it, then you've got to carry the right to be forgotten all the way through those websites.

So, what is the significance of that to us in the United States? Well, if they're compelled to take out information, to take down, to remove, to delist information that you would otherwise be able to find in the United States, it will no longer be available to you. And this is in order to protect the rights of someone in Europe. In the US, we don't have the right. We can't ask Google to take down negative information about us. If there was a copyright violation, we can ask them to do that, but not privacy. But, if you're going to do a search on Google.com, there will not be as much information there as there once was. And so it does impact us in the United States.

Just to close here, I just want to explain if you have something to hide, you can now take steps and then safely say [singing]:

The light is green, the web is clear,
So if you want to go surfing, dear,
I'm delighted, I'm delisted, I'm de-Googled!

I understand the reason why
You're curious, and just want to pry:
You're de-nosey, you're de-snooping, you're
de-peeping

You can tell at a glance
That the EU has taken a stance.

You can hear their Court of Justice
Murmuring low:
"You'll never know."

So, please be sweet, my chickadee,
And when you ask me, I'll say to thee,
It's delightful, it's delisting,
It's debatable, it's deletable,
It's defensive, it's delib'rate, it's deleted,
It's de-Googled.

Lisa Macklin: So, that's a hard act to follow. I do have to tell you that I once sent a friend of mine a birthday card that said, "In honor of your birthday, I will not sing." And so I will not sing and you should be very grateful for that. So, I'm going to talk to you about privacy in libraries and really try and take what Gary and Bill have said and kind of bring them back into our everyday lives, if you will.

So, in the US we have several sources of privacy. As Bill mentioned, in the US we don't have this sense of privacy and privacy laws that they have in Europe, and so we find that we do have a patchwork, if you will, of federal laws—FERPA being an example within higher ed—and state laws. There is a common law right of privacy and there are some state statutes for privacy, and then we also have some state statutes specifically for libraries that really do vary by state, and they do dictate the confidentiality of library records. However, those state laws often don't include library records that might be hosted by a third party, for example. And there's a list of various state laws on the ALA website, so if you're not familiar with your state law I would suggest that you take a look there and really begin to understand what your state law says regarding privacy.

We also have institutional policies on privacy if you're in an academic institution; however, much of our own library policies really are not stemming as much from a law as they are stemming from the ALA code of ethics. And essentially that code of ethics has been the grounding of library policy for a number of years, and it's essentially protecting the library user's right to privacy and to confidentiality. This goes to that core library mission of providing open and equal access to our

collections, and we do this in order to enable education and free inquiry, those kinds of cornerstones of a democracy. However, privacy is not absolute, and we have to balance user privacy with our other constraints around efficiencies, requirements of law enforcement, and then just plain old Internet security. Much of our privacy practices are really rooted in our history of deleting circulation records when somebody has returned a book. They're really grounded in the days of almost complete print collections. However, technology has expanded both our collections and our services, and Gary has demonstrated how easy it is to track a user on the Internet. So, let's talk about doing a privacy audit for library services and what you might want to think about.

Three questions that I think are worth asking . . . First of all, do you have a library privacy policy? Have you reviewed it lately? Does it pertain only to your print collections? Does it apply to online collections and take into account the changes that we've seen recently? Does it align with federal and state law in your institutional policies? I came up with, well, let me backtrack a tad bit. I do want to say that policies are really important because they provide transparency, and they do provide assurances to our users, but it's in implementing practices that I really think we can make the most difference. And I do think that we all have a role to play in that, so I came up with kind of these three questions. Use them if they're useful to you, but I don't think that we always ask these questions when we should.

The first is, what data is being collected? And Gary gave us examples that I don't think some of us are always aware of. Who has access to that data? And then do users have the option to opt out? So, putting that another way, can users search, access, and read anonymously? Or, if they're using a particular service, are they actually automatically getting tracked in ways that they don't realize? For systems librarians and technologists—really looking at ILS systems, search and discovery systems, mobile apps, library webpages—we need to consider what data is really being collected. How securely is it stored? How securely is it transmitted? Is personally

identifiable information encrypted? For our vendor colleagues, have you tested your systems for vulnerabilities? If you provide hosting services, how securely are patron and user data stored? For those who do usability testing—of which, quite frankly, I'm a big fan—I think we need to be aware for collecting personally identifiable information as part of usability testing, realize that we probably need to be anonymizing that information. We need to make certain that the participants understand the parameters of the data collection, and the use that we're going to make of that data. For scholarly communications librarians, most of our institutional repositories do have basic hit and download information, which is pretty generic and anonymized, but we do have some repositories that have sensitive data and other kinds of things where you have to register as a user to actually download the files. So how are we tracking that information, storing it? Do we delete it on a regular basis?

For collections librarians, do you know what data the vendor of those collections and online materials is collecting? Is there digital rights management software attached to those files? How is that data collected used? With whom is it shared? Now, data collection is not, in and of itself, a bad thing. We in libraries need to know how our collections are used. We make collection decisions on a regular basis based on use statistics that are provided from vendors. But do we need that information for each individual user, or is that information in the aggregate really what it is that we need to make those decisions? So, for electronic resources librarians, are you negotiating for terms in your license agreements that protect user confidentiality and privacy? We can do, by contract law, what we don't necessarily in this country have laws to protect privacy for. For publishers and content providers, we need your help. We understand that you have a desire to track your content and how it's used, but we ask that you also understand that tracking and storing who is reading what is in direct conflict with our values and our ethics. Katherine touched on this in her presentation as well.

Liaisons and subject librarians, I will echo what Gary has already said—that we really need to be

teaching our users, including our faculty as well their students, about privacy as part of information and digital literacy. So you have a very big role to play in educating users. We need to support our users' ability to make informed choices. I've used Facebook here as an example, in part because they have improved their user privacy over the last several years, often as a way of reacting to user feedback. And when I logged into Facebook a couple of days ago it asked me if I wanted to review my privacy settings, so they've gotten more proactive in how they really approach privacy. But I think we in libraries really need to be proactive. We shouldn't be waiting for an incident or user demand to really be addressing user privacy.

For all of us, whether you are a librarian, a technologist, a publisher, or some combination of the three, you need to understand the contours of privacy law, and our legal and ethical obligations. And most importantly, as Gary showed, we really need to understand how technology works, both to invade our privacy and to protect our privacy. We need to regularly question privacy implications of new tools and services. And we need to keep current, which I realize is a huge

challenge since technology is going to continue to change. So Gary's infoDOCKET is one great example of that. The ALA's Washington office is also another example where they regularly will have information about privacy.

Finally, and most importantly, be an advocate. We all have a role to play, and I've given you this laundry list not expecting you to go back to your institutions and do everything all at once. But my hope is that each of you can have a takeaway from this panel discussion to do, that you will go back to your institutions and think about and consider. As our collections are becoming increasingly digital, and our services are often hosted on platforms and servers even outside of our own institutions, our sensitivity to patron privacy really does need to increase. Fulfilling our library mission of providing equal and open access to our collections to enable education and to enable free inquiry relies on libraries remaining a trusted institution. When users recognize or fear that their privacy or confidentiality is compromised, then the freedom of inquiry no longer really exists. And that, I think, is the thing we need to guard against. Thank you.

Reference

European Commission. (2015, June 15). *Stronger data protection rules for Europe* [Press Release Database; European Commission—Fact Sheet]. Retrieved from http://europa.eu/rapid/press-release_MEMO-15-5170_en.htm