COMMENTARY

Choose Privacy Week 2018

Big Data Is Watching You

Author _ William Marden (williammarden@nypl.org), Chair,
ALA Intellectual Freedom Committee's Privacy Subcommittee

[EDITOR'S NOTE: This introduction and the six articles that follow are from the American Library Association's 2018 online symposium for Choose Privacy Week, which began on April 16, 2018.]

Five months ago, when the members of ALA's Privacy Subcommittee met to decide on this year's [2018] "Choose Privacy Week" (CPW) theme, it's a fair bet to say that only a tiny percentage of the general public had ever heard of Cambridge Analytica, Aleksandr Kogan, the SCL Group, or of a fairly obscure app called "thisisyourdigitallife."

And yet, there were warnings about Cambridge Analytica's program as early as December 2015, when the London *Guardian* first reported on this data-collection program and its integration with Facebook as part of Ted Cruz's 2016 bid for the US presidency. Michael Zimmer, a University of Wisconsin-Milwaukee associate professor and a member of ALA's Privacy Subcommittee, was quoted by the *Guardian* about why the use of such data was highly problematic. "It's one thing for a marketer to try to predict if people like Coke or Pepsi," said Zimmer, "but it's another thing for them to predict things that are much more central to our identity, and what's more personal in how I interact with the world in terms of social and cultural issues?"

In the wake of Mark Zuckerberg's Congressional testimony last week [in April 2018] and the related explosion of public interest in how online personal data is collected, stored, shared, used, and sometimes misused, this year's CPW theme—"Big Data is Watching You"—could not be more perfectly timed.

In the library community especially, the right of library users to keep private their use of library resources has traditionally long been a hallmark of the ALA's principles, embedded in its "Library Bill of Rights" and actively promoted by ALA's Intellectual Freedom Committee and its Privacy Subcommittee.

But those rights and protections are increasingly being challenged by the use of "big data": library patron information that is bundled up, aggregated and usually (but not always) anonymized for varied purposes including trend analyses, grant funding, and reporting to local governments.

Has this new era of data collection become another form of surveillance? Is the aggregated data of library users truly anonymous? Can we collect such data and still guarantee the minimum standards of privacy for our library users?

This is the theme explored by the authors featured in the 2018 online symposium for Choose Privacy Week. Each writer explores the issues around Big Data, and promotes methods and technologies to help guide librarians in knowing how to responsibly use these data-gathering techniques.

Libraries as Private Spaces

Author _ Jason Griffey (griffey@gmail.com), affiliate fellow, Berkman Klein Center for Internet and Society

The modern world is largely driven by what has been termed "surveillance capitalism" by Shoshana Zuboff. Surveillance capitalism is the monetization of data that is gathered through the observation of individual or group behavior. This data can be gathered voluntarily (by asking users for it), involuntarily (one company gathering information about an individual by taking it from another data source), or via some combination of the two (data that was given freely by the individual, but is later leaked or stolen from the recipient). Almost the entirety of the modern web is predicated on surveillance capitalism, with targeted advertising being the driving force behind many of the largest companies in the world. Nearly every social network (Facebook, Snapchat, Instagram, and the like) are in this category, as are the largest web retailers like Amazon. Google is, famously, not really a search company, nor is it driven by a desire to organize the world's information. It is an advertising company, with 90 percent of its revenue coming from some form of advertising based on the things it knows about you.

Consider, just as an exercise, how much Google can know about you. If you use the Google search engine, it knows everything you've searched for, every result, and every link you've clicked to get information. If you use the Chrome browser, then Google has the capacity to know nearly everything. In theory, they can know everything you've typed into the address bar, everything you've typed into a non-secure form, and more. If you use Gmail, Google scans your email (sent and received) to better target you. Use Google Drive or Google Docs/Sheets/Slides, and those are scanned as well. If you or your ISP use Google's DNS service, they gather information about what site is requested, what geographic area you are in, and more.

The last few months have brought to light the cost to society of surveillance capitalism, in the form of Facebook and the potential influencing of the US elections through automated targeted advertising. There is beginning to be a backlash against this type of data collection, and it's possible that the near future may see the rise of regulation and policy to prevent this sort of data from being used in advertising. This isn't out of the realm of possibility, as the US has a history of federally regulating types of advertising allowed, from type (subliminal advertising) to content (cigarette ads, alcohol ads).

This is likely to be necessary as future hardware developments allow for near-zero-cost low power data collection systems to be implemented ubiquitously throughout our world. Consider the development of a camera module that powers itself, because the sensor is also a solar cell that produces the power necessary to run itself. Due to Moore's Law and Koomey's Law we will soon have the capacity to spread cameras and microphones with cellular and Wi-Fi radios attached to them across our environment at incredibly low costs. It is very easy to imagine a future where companies like Google give away packages of these "ubiquity sensors" and use them to harvest data about movement and behavior in the same kind of way that they "give away" Google Maps by harvesting movement information from Android phones.

Once we head further down this road, it is highly possible that we are approaching the end of public spaces being anonymous or private spaces where one can be reasonably certain they aren't being surveilled. Right now, this is already the case in many cities like London, and we have seen omnipresent surveillance spread across entire counties in the case of somewhere like China. As these chips get smaller and more energy efficient, we approach a sort of sci-fi "smart sand" which can be sprinkled across spaces in order to gather data of nearly any sort, and the cost of which will be a rounding error in the budgets of major cities. The combination of private commercial interests and government surveillance will quickly render every available square foot of populated areas a target of surveillance.

This world of smart sand is, make no mistake, a dystopia. Regardless of the good that is possible in such a world (no one would ever collapse on the sidewalk without an ambulance being called, because the sidewalk itself would call . . .), the ultimate state is not a good one. The removal of the expectation of private actions and speech is the strongest possible type of prior restraint for democratic action. I believe that not only are privacy and security fundamental to the operation of libraries in the United States, but soon, libraries may be the last public space that doesn't surveil you for the purposes of increasing the bottom line of a corporation. This is a space and effort that libraries should embrace, advertise, and focus on…privacy and freedom from surveillance is necessary for a functional democracy, in the same way that the ALA's Democracy Statement says:

> "Democracies need libraries. An informed public constitutes the very foundation of a democracy; after all, democracies are about discourse—discourse among the people. If a free society is to survive, it must ensure the preservation of its records and provide free and open access to this information to all its citizens."[i]

The power and strength of the library to protect and enable democracy and equity go farther than the preservation and access to information. Libraries have a duty to the privacy of their patrons, but moreover we have a duty to defend the foundations of democracy itself. In this future world of ubiquitous surveillance, the library has a duty to say no, and to draw a hard line against the rise of ubiquitous technological surveillance. Libraries are spaces where people should be safe, as safety is a prerequisite for information seeking and understanding. Ubiquitous surveillance is fundamentally unsafe for vulnerable populations of patrons, and libraries have a duty to those patrons to resist the collection and retention of data about individuals.

As a result, the next five to ten years are going to be incredibly dangerous. Libraries can step up at every level to protect the privacy and security of their community. In order to protect and support the fundamental tenets of our democratic society, libraries must double-down on privacy now by protecting their patron's data and information seeking and must also be ready to protect their communities by resisting the rise of ubiquitous surveillance in the world.

[i] "Democracy Statement," http://www.ala.org/aboutala/governance/officers/past/kranich/demo/statement (accessed on December 10, 2018).

Big Brother Is Watching You

The ethical role of libraries and big data

Author _ Erin Berman, (erinberman@aclibrary.org), innovations manager for the San Jose Public Library

The Ethics of Privacy in Librarianship

Libraries are one of the most trusted institutions in our country. People place librarians in the same class as doctors, nurses, firefighters, and teachers. Communities bemoan the possibility of their local libraries closing, with two-thirds saying it would have a major impact on their lives if the library doors were shuttered.[ii]

One of the key reasons libraries are held in such high regard is because we operate under a code of ethics; a code that is driven by intellectual freedom and ensures the public has freedom of access to information.

Although that code of ethics may not be at the forefront of your mind day-to-day, it is the backbone of our institutions. Our code of ethics is the foundation upon which our libraries were built and the reason that we are such a trusted part of every community. It is our responsibility as library professionals to use this code as a guide to drive institutional operations and as a "north star" for ethical dilemmas.

[ii] "Americans' attitudes toward public libraries," John B. Horrigan, Pew Research Center: Internet and Technology, http://www.pewinternet.org/2016/09/09/americans-attitudes-toward-public-libraries/ (accessed on December 10, 2018).

We find ourselves at a precipice, faced with a huge ethical decision about how libraries will interface with the privacy of our patrons. The current code of ethics states that we will, “Protect each library user’s right to privacy and confidentiality with respect to information sought or received and resources consulted, borrowed, acquired or transmitted.”[iii] Libraries committed themselves to this ethical charge nearly eighty years ago, becoming known around the world for our privacy advocacy, championing patrons’ rights to access information without scrutiny.

Our New, Happy life

The world is very different today than it was eighty years ago. Today, the wealth of human knowledge is available in our pockets. The media mourns the lost age of privacy, telling us it is an outdated concept; dead and buried. People freely part with their personally identifiable information, giving companies the ability to create detailed profiles of their lives to sell to the highest bidder. Our information has become a commodity, bought and sold to provide us customized services and sell us more things.

We have entered the age of big data; it is virtually impossible to move through this world without being tracked, labeled, and categorized. Algorithms sort us into consumer categories that are used to sell targeted advertising and provide different levels of services based on who they think we are. If you would like a glimpse into how you are labeled and categorized, take a look at your Facebook ad profile.

Big data analytics are all about gathering as much information as possible in order to predict the future. By gathering data about how you use a company’s product, companies can make adjustments to services on an individual level. This has led to some amazing online experiences! I love my tailored suggestions on Netflix and if I have to see advertising I would much rather it be for something I might actually purchase. Most people have no concept of the amount of information they are giving away but do reap the rewards of personalized internet experiences.

Big Data in Libraries

So, where does that leave libraries? Never wanting to be left behind, libraries are finding themselves pushed to participate in the world of big data analytics. Companies offer libraries an unprecedented glimpse into how our patrons use library services. These products allow libraries to track an individual patron’s interaction with all aspects of the library. Collected data sets may include:

- All ILS [Integrated Library System] data (e.g., name, age, address, email, phone, driver’s license, gender)
- Borrowing history
- Program attendance
- Website interactions

In some instances, library data is then paired with data from credit reporting agencies. Libraries can see maps that give household level information about patrons based on the category in which the company’s algorithm has placed them. This includes:

- Income
- Number of children in the house
- How long they have lived at their residence
- Spending habits
- Hobbies
- Device usage
- Media consumption preferences
- And much more!

These companies are telling libraries that our patrons are demanding personalized services, that we are facing a future of irrelevance. Luckily for us, they say, their products have all the answers. By tracking patron behavior, we can give them the experience they have come to expect from this new digital world. Libraries can segment out our patrons, sending targeted marketing based on their behaviors, customizing our services based on what they read and what programs they attend. We will finally be able to use real data to tell our stakeholders why we are of value, so they won’t withdraw our funding. This messaging is a classic anxiety stick, followed by a marketing carrot.[iv]

For a more detailed look at the data you can access by using these companies’ products, take a look at some of the big ones on the market right now:

- OrangeBoy’s Savannah
- Gale’s Analytics on Demand
- OCLC’s Wise

[iii] ALA Code of Ethics, American Library Association, http://www.ala.org/tools/ethics (accessed on December 10, 2018).

[iv] “Lessons from the Facebook Fiasco,” Barbara Fister, *Inside Higher Education*, April 15, 2018, https://www.insidehighered.com/blogs/library-babel-fish/lessons-facebook-fiasco (accessed December 10, 2018).

Our Ethical Responsibility

Libraries are often the only access point to information for the most vulnerable members of our communities. We welcome our undocumented citizens, those living unhoused, mentally ill, and minorities of all kinds. We have a responsibility to all of our patrons to protect and fight for their privacy. Our patrons trust us because they know that when they walk through our doors, we are there to help them access information needed to become their best possible selves. We don't judge them based on who they are or why they are there. Our doors are open to all.

How many years have libraries been told that they are on the brink of destruction? Adapt or die. Libraries have made amazing changes in our service models by shifting our concept of literacy and extending its reach into technology and providing hands-on programming. Those changes have remained rooted in the core fundamentals of librarianship. On the other hand, the adoption of big data analytics that allows us access to patron-level data about their use of library services goes against one of our most valued core ethics.

It is our ethical responsibility to ensure that all of our patrons have the "right to open inquiry without having the subject of one's interest examined or scrutinized by others."[v] This includes us as library workers. We do not have the right to examine or scrutinize what our patrons do in the library, making decisions on how to treat them based on those behaviors. Even something as seemingly harmless as sending targeted marketing emails means that you are judging who a patron is and what their future behaviors may be based on their reading habits and library usage.

If the recent news surrounding Facebook can be any kind of lesson to us, it is that privacy is not dead in the minds of our patrons. People are seeking control over their information, as evidenced by the sweeping new privacy regulations that went into effect in the European Union in May 2018. Consumers want to use services they trust. We already have that trust; why take steps to erode our standing in regards to the protection of our patrons' privacy?

Moving Forward

Let us flip the narrative we are being sold by the big data analytics companies. Now is the time for us to tout the virtues of the library as a privacy haven to our patrons. We are not Amazon, Barnes and Noble, or Google, and we should never strive to be. Our patrons are not our products. That is a huge difference between public institutions like libraries and private industries like social networks and tech conglomerates who derive their earnings from advertising.

Libraries are the cornerstone of democracy. We have a democratic duty to uphold the privacy ethics of librarianship and not track and allow third-party access to our patrons' information. Remember, once those datasets are created anyone can gain access to them. Do you want our government having these detailed reports on our patrons? If not, then it is time to rethink how we move forward in doing business with these companies.

Do not jump into big data without being intentional, transparent, and having a comprehensive understanding of how the products work. Utilizing different datasets to drive decision making and analyze the work done in libraries is extremely important, but it must be done with careful attention paid towards protecting our patrons' privacy. When moving forward with a big data contract, consider these guidelines for use:

- Only collect data in the aggregate or anonymously. Do not collect personally identifiable information (name, email, address). Do not track any reading or library usage data on specific patrons.
- Transaction-level data that uniquely identifies both a patron and an item should be avoided unless required for a specific and limited purpose.
- Patron consent must be gained to collect any transaction-level data that links a patron to an activity (e.g., books read, programs attended, e-resources used).
- Vendors must have a public privacy policy on their website that adheres to industry standards.

It is up to us as library professionals to shape the future of our institutions. Will we continue to uphold the ethics of our profession, ensuring that we remain a trusted source of information for our citizens? Now is the time to act! Let's explain to our patrons what sets us apart and remind them that we will continue to be their privacy advocates and champions.

[v] "Privacy: An Interpretation of the Library Bill of Rights," American Library Association, http://www.ala.org/advocacy/intfreedom/librarybill/interpretations/privacy (accessed on December 10, 2018).

The Challenge of Balancing Customer Service with Privacy

Author _ **Sarah Houghton** (librarianinblack@gmail.com), director of the San Rafael Public Library

The third principle of the American Library Association Code of Ethics is that "We protect each library user's right to privacy and confidentiality with respect to information sought or received and resources consulted, borrowed, acquired or transmitted." That sentence leaves no room for misunderstanding.

On the whole, the history of libraries is one of service and safeguards against censorship, inequity, and privacy violations. Despite librarians' long-held professional ethics, the reality is that the advent of the digital data brought privacy concerns from day one—for the few who were paying attention. The one-on-one inviolable relationship between library worker and library customer fundamentally changed with digital data. What used to be a sacrosanct, trust-based bond between two parties turned into a cloudy series of agreements and contracts that, frankly, were not fully understood by the libraries that entered into them or the customers who clicked the "I agree to the terms of service" button. Suddenly, third parties such as integrated library systems, scholarly databases, and eBook vendors had access to library patron databases and all they held—personally identifiable information including social security numbers, birth dates, and driver's license numbers, along with the standard name, address, and other contact information.

Libraries then entered a race to catch up to the technology with our policies and practices, but the tech moved faster than we did. Inconsistently, but steadily, libraries started removing non-essential data from our patron databases and reevaluating contracts with digital providers. Concurrently, companies that libraries did and do business with used cookies to track the activities of library users within their products, then used that information to market services to those people. Companies made arrangements with fourth parties to connect library customer activity in a product with their personal, non-library use. Companies sold aggregate data about library customer activities. And all the while libraries and librarians have been unaware or unwilling to confront the violations of one of our core values. The sad truth is that most libraries have no idea what data vendors are collecting on our customers or simply do not care enough to prioritize customer privacy over concepts like ease of access, provision of digital content, and user demand.

People have developed an intensely complicated relationship with technology and privacy. On one hand, technology and digital data have made it easier to provide personalized online experiences. On the other hand, people are often surprised to discover how much of their privacy they have traded for those personalized experiences. How do we, as libraries, find that balance between customer service and privacy?

Enter the world of big data. Companies read the library marketplace and saw a space for data analytics in library services with a noble goal in mind—to create enough trend data to lead to data-based decision-making. With the tools at our disposal we can now, with a few clicks and search terms, bring up a map that will, in essence, show you or me that someone at a particular address checked out a particular book in the last week. I don't want that level of information to be available to me, to a third party vendor, or to anyone else. We should all be disturbed by the level of specificity and personally identifiable information used by the big data companies in the library marketplace.

I wither as I see more and more libraries increasingly using data collection (that would have been unheard of in past decades) for tracking customer usage, analyzing trends in use, creating fancy looking reports for their parent agencies, and storing and sharing data in ways that are increasingly hackable and shareable.

This is not a problem that is solely ours. Every organization or individual that collects data about the activities and profiles of people is facing this same conundrum. This seems like a natural place for libraries to take the lead in big data and user privacy. To draw a line in the sand and say no further. To date, we have not done that collectively or individually (for the most part). One of my greatest professional regrets is prioritizing what my customers and stakeholders say they want over my own understanding of the stringent privacy and confidentiality practices that I should be honoring as a librarian. I, like most of my peers, give the community what they want at the expense of their ability to control their personal data in an informed and conscious way.

Just because technology makes something possible doesn't mean it's something we should do. In most cases, decision makers and customers alike are unaware of the potential privacy issues with their data until that data is exploited by others. We are one big library data breach away from this issue becoming nationwide kitchen table conversation.

So how do we balance the potential of big data and privacy and confidentiality? A "scorch the earth" policy seems the most logical in my mind—for libraries to cease keeping any non-essential data and refusing to do business with any company that does otherwise. But I am also a realist and know that the thousands of libraries in the United States alone are not going to be able to come to agreement on that level of stringency.

I do think, however, there are things that all libraries can do to better uphold our values. Make data policies and practices transparent. Any collected data should have a clear purpose. Ensure data quality and deletion when no longer necessary. Renegotiate contracts to ensure greater transparency, authentication, back-up, replication protections, and security protocols. Provide clear information on how customer data is going to be used. Provide a mechanism for an individual to review their personal data on our systems.

And above all, keep repeating that mantra—the third principle of the American Library Association Code of Ethics—and remember that privacy protection is part and parcel of what we signed up to do as librarians.

Practical Privacy

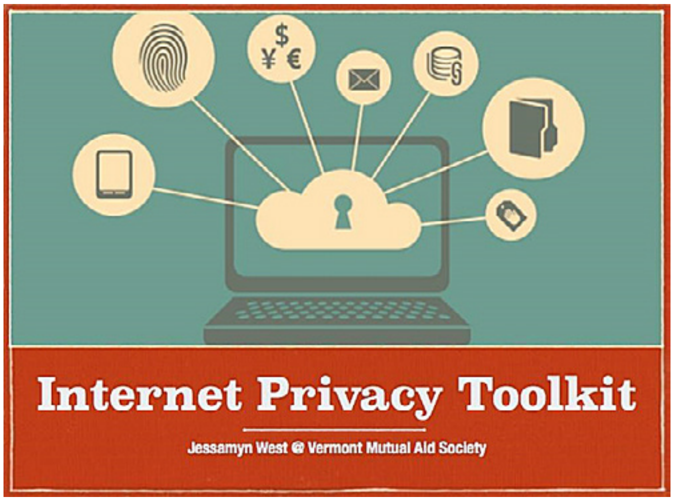
Helping People Make Realistic Privacy Choices for Their Real Lives

Author _ Jessamyn West (jessamyn@gmail.com), community technologist and librarian, Vermont Mutual Aid Society

Getting good information about how to secure your privacy online can be challenging. This is doubly true if you are a technology novice, struggling to keep track of your own passwords, devices, and secret questions. You have the dual-pronged concerns of needing to learn this stuff from someone, but also not being sure whom you should trust. And many people are telling you the situation is urgent but then leaving you on your own to fix it. I've worked on an approach I feel is useful.

I run a Drop-In Time for the people in my rural Vermont community. For thirteen years people have come in with their technology questions and we've muddled through figuring technology stuff out together. Originally there were a lot of people needing mouse skills, email accounts, and assistance setting up Facebook. Now, there are more people with general questions about privacy, learning about the cloud, how material gets shared and re-shared, and, again, Facebook.

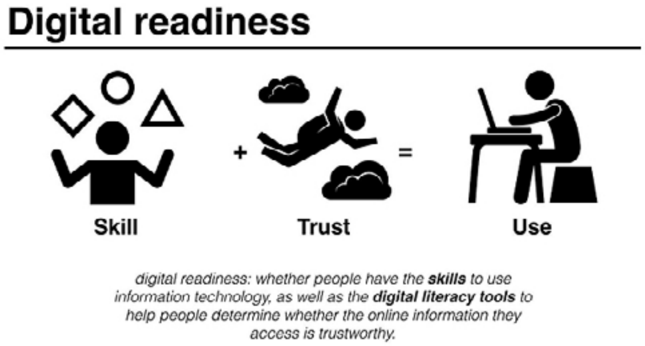
The Pew Center for Internet and Society, one of the few larger non-academic organizations continuing to do research into the digital divide, has been talking about digital literacy and the corresponding digital readiness gap. Whereas the digital divide used to refer to a lack of actual computers or sufficient broadband, now we're seeing more people struggling with their inability to engage in self-directed learning online, while gaps in technology and broadband access still remain.



Internet Privacy Toolkit

Jessamyn West @ Vermont Mutual Aid Society

This challenge—coming at a time when more and more learning opportunities are provided via technology and not an in-person human facilitator—risks leaving the most vulnerable people further behind. Pew's research indicates that to be able to be digitally ready, users must not only have basic technology access and skills, but also an ability to evaluate and discern trustworthiness of content online, and to trust their own judgment in the absence of someone to ask. Here's a slide I made for a talk I frequently give about how libraries can bridge the digital divide.



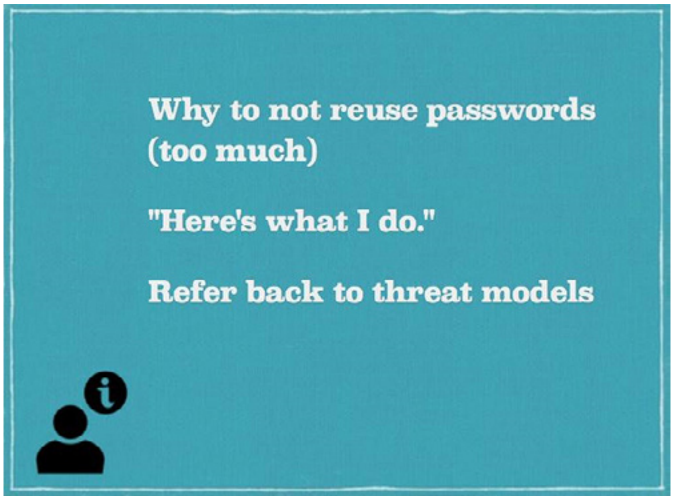
Digital readiness

digital readiness: whether people have the ***skills*** *to use information technology, as well as the* ***digital literacy tools*** *to help people determine whether the online information they access is trustworthy.*

So part of the new role of librarians is helping people work on evaluating online information and learning what to trust and what not to trust. This helps our patrons be better prepared for online learning, so they can teach themselves the things they would like to learn. I figured it would be useful to help bring a privacy-oriented version of Drop-In Time to libraries. Starting with a talk at my own local library, I created a very simple slide deck with a few basic privacy topics, intended more as a set of conversation starters than an all-inclusive list. I added a list of links, also available as a Google document, where people could go for more information on the specific topics we mentioned. A short list of topics with some sample sub-topics is here:

- **Threat modeling** – Before you decide what to do about online privacy, it's important to look at your own personal situation and think, "How at-risk am I? Where's my personal line between convenience and privacy?"
- **Passwords** – How to choose a good one. Why the passwords that websites make us choose are so complicated. Sensible talk about whether to write them down or not.
- **Internet Traffic** – What to know about using public WiFi. How much does your browser know about you? What are some more secure options for browsing? What browser plug-ins are helpful?
- **Listening/Recording** – What is the Internet of Things all about? Should you cover your laptop's camera? What about all those "intelligent personal assistants" people use nowadays?
- **Tracking** – What are cookies? How do these advertisers seem to know so much about me? What are strategies for not being tracked online?
- **More information** – Who are good people to listen to? Who should I not listen to?

I learned a lot by giving this talk in many public libraries around the state of Vermont. Even people who were fairly sophisticated about technology still had gaps in their knowledge. This is not surprising, considering that the business model of many popular online services is, essentially, privacy violation, and they spend millions of dollars to obscure the fact that this is what they are doing. Many people also had different ideas of what "best practices" were, and enjoyed sharing what they knew with other patrons in a guided conversation. It was helpful for people who were anxious about their own technology usage to know that it fell within "normal" boundaries, and that many other people also had questions about the same topics.



Why to not reuse passwords (too much)

"Here's what I do."

Refer back to threat models

I eventually adapted this talk to be a talk *about* the talk, which I presented at the New England Library Association conference last year. It was intended for librarians who wanted to give similar talks at their own libraries. It included interstitial slides with specific advice for librarians about how to present the topics, as well as strategies for how to set up technology in their own libraries. I encourage you to use it if it solves a problem for you.

My general point is this: you don't have to be a privacy expert in order to help people learn to protect their privacy online, and sometimes it helps if you aren't. People are currently getting a lot of bad information from businesses that are trying to either sell something to them or encourage their online oversharing so that they can sell their data. Some people may not mind this, but many others certainly do. People are eager for straight talk from trusted people, presented in a non-judgmental fashion; so much tech advice online nowadays takes the form of "You are an idiot if you aren't doing things like I do." If we want to give people information in a way that they will understand it and take it to heart, we need to have conversations with them, not give them a list of rigid rules.

The librarians' position in society as a trusted purveyor of solid information puts us in a fairly unique role as people who can and should be front and center of online privacy discussions. Helping people understand privacy helps them make better choices more tailored to their own lives and the information needs they encounter.

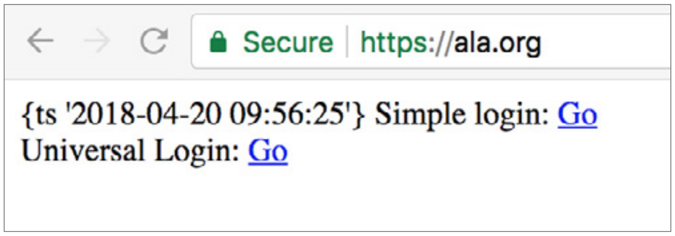
Your Library Organization Is Watching You

Authors _ Eric Hellman (eric@hellman.net), president of the Free E-book Foundation and founder of unglue.it, and **T.J. Lamanna** (professionalirritant@riseup.net), chair of the New Jersey Library Association Intellectual Freedom Committee and the emerging technologies librarian at the Cherry Hill Public Library

We commonly hear that "Big Brother" is watching you in the context of digital and analog surveillance such as Facebook advertising, street cameras, E-Z pass highway tracking, or content sniffing by internet service providers. But it's not only Big Brother; there are a lot of smaller, less obvious "Little Brothers" as well, that wittingly or unwittingly funnel data, including personal identifiable information (PII), to massive databases. Unfortunately, libraries (and related organizations) are a part of this surveillance environment. In the following article, we'll break down two example library organization websites. We'll be focusing on two American Library Association (ALA) websites: ALA's Office for Intellectual Freedom's *Choose Privacy Week* website (ChoosePrivacyWeek.org) and ALA's umbrella site (ala.org).

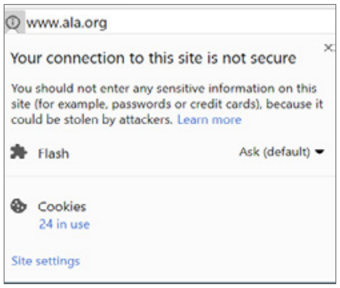
Before we dive too deeply, let's review some basics about the data streams generated by a visit to a website. When you visit a website, your browser software—Chrome, Firefox, Safari, etc.—sends a request containing your IP address, the address of the web page you want, and a whole bunch of other information. If the website supports "SSL," most of that information is encrypted. If it does not support "SSL," it is not encrypted, and network providers are free to see everything sent or received.

Without SSL, bad actors who share the networks can insert code or other content into the web page you receive. The easiest way to see if a site has a valid SSL certificate is to look at the protocol identifier of a URL. If it's 'HTTPS', that traffic is encrypted; if it's 'HTTP,' DO NOT SEND any personally identifiable information (PII), as there is no guarantee that traffic is being protected. If you're curious about the quality of a site's encryption, you can check its "Qualys report," offered by SSL Labs, which checks the website's configuration, and assigns a letter grade. ALA.org gets a B; ChoosePrivacyWeek gets an A. The good news is that even ALA.org's B is an acceptable grade. The bad news is that the B grade is for "https://www.ala.org/", whose response is reproduced here in its entirety:



{ts '2018-04-20 09:56:25'} Simple login: Go
Universal Login: Go

Unfortunately, the ALA website is mostly available only without SSL encryption. You don't have to check the SSL Labs to see the difference. You can recognize ChoosePrivacyWeek.org as a "secure" connection by looking for the lock badge in your browser; click on that badge for more info. Here's what this looks like in Chrome:



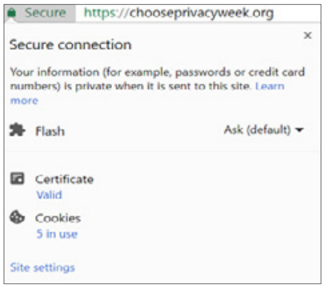
Your connection to this site is not secure

You should not enter any sensitive information on this site (for example, passwords or credit cards), because it could be stolen by attackers. Learn more

Flash Ask (default)

Cookies
24 in use

Site settings



Secure connection

Your information (for example, passwords or credit card numbers) is private when it is sent to this site. Learn more

Flash Ask (default)

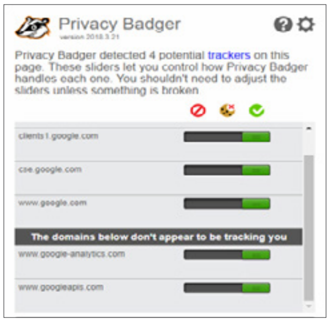
Certificate
Valid

Cookies
5 in use

Site settings

Don't assume that your privacy is protected just because a site has a lock badge, because the web is designed to spew data about you in many ways. Remember that "whole bunch of other information" we glossed over above? Included in that "other information" are "cookies" which allow web servers to keep track of your browsing session. It's almost impossible to use the web these days without sending these cookies. But many websites include third party services that track your session as well. These are more insidious, because they give you an identifier that joins your activity across multiple websites. The combination of data from thousands of websites often gives away your identity, which then can be used in ways you have no control over.

Privacy Badger is a browser extension created by the Electronic Frontier Foundation (EFF) which monitors the embedded code in websites that may be tracking your web traffic. You can see a side-by-side comparison of ALA.org on the left and ChoosePrivacyWeek.org on the right:

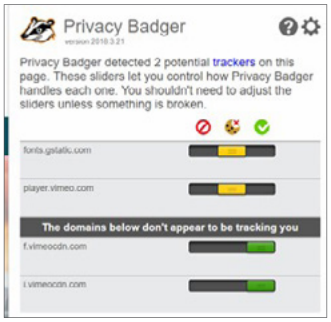


Privacy Badger detected 4 potential trackers on this page. These sliders let you control how Privacy Badger handles each one. You shouldn't need to adjust the sliders unless something is broken

clients1.google.com
cse.google.com
www.google.com

The domains below don't appear to be tracking you

www.google-analytics.com
www.googleapis.com



Privacy Badger detected 2 potential trackers on this page. These sliders let you control how Privacy Badger handles each one. You shouldn't need to adjust the sliders unless something is broken.

fonts.gstatic.com
player.vimeo.com

The domains below don't appear to be tracking you

f.vimeocdn.com
i.vimeocdn.com

The 2 potential trackers identified by Privacy Badger on ChoosePrivacyWeek.org are third party services: fonts from Google and an embedded video player from Vimeo. These are possibly tracking users but are not optimized to do so. The 4 trackers on ALA.org merit a closer look. They're all from Google; the ones of concern are placed by Google Analytics. One of us has written about how Google analytics can be configured to respect user privacy, if you trust Google's assurances. To its credit ALA.org has turned on the "anonymize IP setting," which in theory obscures user's identity. But it also has "demographics" turned on, which causes an advertising (cross-domain) cookie to be set for users of ALA.org, and Google's advertising arm is free to use ALA.org user data to target advertising (which is how Google makes money). PrivacyBadger allows you to disable any or all of these trackers and potential trackers (though doing so can break some websites).

Apart from controlling the giving of data to third parties, any organization must have internal policies and protocols for handling the reams of data generated by website users. It's easy to forget that server logs may grow to contain hundreds of gigabytes or more of data that can be traced back to individual users. We asked ALA about their log retention policies with privacy in mind. ALA was kind enough to respond:

"We always support privacy, so internal meetings are occurring to determine how to make sure that we comply with all applicable laws while always protecting member/customer data from exposure. Currently, ALA is taking a serious look at collection and retention considering the General Data Protection Regulation (GDPR) *EU 2016/679,* a European Union law on data protection and privacy for all individuals within the EU. It applies to all sites/businesses that collect personal data regardless of location."

Reading in between the lines, it sounds like ALA does not yet have log retention policies or protocols. It's encouraging that these items are on the agenda, but disappointing that it's 2018 [at the time of writing] and these items are on the agenda. ALA.org has a 4-year-old privacy policy on its website that talks about the data it collects, but has no mention of a retention policy or of third party service use.

The ChoosePrivacyWeek.org website has a privacy statement that's more emphatic: "We will collect no personal information about you when you visit our website unless you choose to provide that information to us."

The lack of tracking on the site is aligned with this statement, but we'd still like to see a statement about log retention. ChoosePrivacyWeek.org is hosted on a Dreamhost WordPress server, and usage log files at Dreamhost were recently sought by the Department of Justice in the Disruptj20.org case.

Organizations express their priorities and values in their actions. ALA's stance toward implementing HTTPS will be familiar to many librarians; limited IT resources get deployed according competing priorities. In the case of ALA, a sorely needed website redesign was deemed more important to the organization than providing incremental security and privacy to website users by implementing HTTPS. Similarly, the demographic information provided by Google's advertising tracker was valued more than member privacy (assuming ALA is aware of the trade-off). The ChoosePrivacyWeek.org website has a different set of values and objectives, and thus has made some different choices.[vi]

In implementing their websites and services, libraries make many choices that impact on user privacy. We want librarians, library administrators, library technology staff, and library vendors to be aware of the choices they are making and aware of the values they are expressing on behalf of an organization or of a library. We hope that they will CHOOSE PRIVACY.

[vi] "SSL Report: ala.org," Qualys SSL Labs, April 10, 2019, https://www.ssllabs.com/ssltest/analyze.html?d=ala.org; "SSL Report: chooseprivacyeveryday.org," Qualys SSL Labs, April 10, 2019, https://www.ssllabs.com/ssltest/analyze.html?d=chooseprivacyeveryday.org.

Patron Privacy and Data Storage

Author _ **Matt Beckstrom** (mbeckstrom@lclibrary.org), systems librarian at the Lewis and Clark Library in Helena, Montana

With all the concerns we have recently regarding privacy and patron information, we sometimes forget about the data we collect and how we store it. This is especially important as we consider all the different ways our data is used. Let us first take a few minutes to look at data storage: what we are storing, where we are storing it, who has access to it, and how long we are keeping it.

During a library privacy audit, it is a good idea to reevaluate what information we are storing. There are many places where we collect information that we may not always think about. Take, for example, our websites. Many web servers by default collect logs containing a lot of information about our users. Not all of it is useful or necessary. It is common to collect information about our visitors' browsers in order to make our websites more efficient—things like browser type and version, operating system, and location. Most of this information is not personally identifiable information and is therefore relatively safe to collect and store, but it is possible to collect too much information. Take a few minutes to verify what information your web analytics is storing, and decide what is the minimum that you need. Many websites will use cookies to track user behavior. If you are using cookies in your website, make sure, if you can, that there is a notice that your website uses cookies. It would be a good idea to provide information to your users on how to block cookies if they do not want to provide this information.

Let us also look at other types of data storage inside the library. For example, we might consider computer usage records or other types of in-house use. Most of us offer computers or other technology for patrons to use, and we need to consider what types of information we store and how long we store it. At my library, we purge patron identifiable information every day. We still keep a record of computer usage, but there is no identifiable information stored with it. For other types of in-house use, like faxes, scanning, or microfilm, we do not store any information about their usage. My ILS does not store a history of patron checkouts beyond two renewals. We purge older financial transactions as well. We keep the number of transactions and the amounts, but we remove any identifiable information from them.

It becomes difficult when we consider third-party companies. There are many systems that connect to our patron databases. For example, downloadable media services connect to our databases using sip2 or APIs in order to authenticate our patrons. During these connections, it is possible that much of our patron information is also being exchanged. When signing up for these services, review their privacy policies. They should cover what types of information they collect, whether it is personally identifiable or non-personally identifiable information, what they use it for, and how long they retain it. For example, the Overdrive privacy policy explains the difference between personal information and non-personal information, and that they only collect non-personal information. It also states that any information they collect is protected and encrypted and is only obtainable by specific employees. They also say that the information is stored for as long as they deem necessary to provide the services they provide, or for as long as is permitted by law. It is also useful to know of any services that your third-party vendors are using. Some companies use other companies for their services. Overdrive, for example, uses Google Analytics and applications like CrazyEgg and Google AdWords. Each of these companies have their own separate privacy policies.

While we may not always have control over the privacy policies of our third-party vendors, we can minimize their access to our information. When they request access to our databases, restrict their access to the smallest amount they need. Purge patron identifiable information before it is provided to third-party vendors or is stored. When possible, negotiate contracts with third-party vendors to minimize the amount of information they collect and how long they store it.

Once we understand all the places where patron information is stored, who has access to it, and how long it is retained, we must provide this to our patrons. We should update our policies to reflect data storage and third-party access to it. In the situation where third-party companies have access to our patron data, we should supply links to their privacy policies. Make it easy for our patrons to know how we use their information, and what they can do to have more control over it. Teach them to understand how to control their information by using privacy protection in their browsers or by reading privacy policies.