Physical-Equivalent Privacy and the Ever Changing yet Same Privacy Landscape:

Challenging Circumstances and Possible Paths for Consideration

Author Note: [a] Presenter, [b] Moderator

[a] Doreen Bradley; ORCID https://orcid.org/0000-0002-1762-0842

2178 Shapiro Library, University of Michigan, Ann Arbor, MI, 48109

dbradley@umich.edu

[a] Dorothea Salo; ORCID https://orcid.org/0000-0002-6388-0311

[a] Gabriel J. Gardner; ORCID https://orcid.org/0000-0002-9996-5587

[a] Lisa Janicke Hinchliffe; ORCID https://orcid.org/0000-0002-5129-4235

[b] Heather Ruland Staines; ORCID https://orcid.org/0000-0003-3876-1182

Abstract

Privacy should be a factor in every decision made around library services and the provision of resources. With staffing stretched to the breaking point, and only more so as a result of the pandemic, what strategies and tools could best help libraries ensure good privacy practices? This session will draw upon current scholarship, grant projects, and policy making in librarianship to highlight ways that librarians are engaging with these important issues. The panel features editors and authors from a special issue of Serials Librarian focused on library privacy. Highlights include discussing physical equivalent privacy, third-party tracking systems, and the development of model privacy license language for vendor negotiations. In this session, participants will talk about the latest concerns and what you might do in your library to assure your faculty, staff, and students that their privacy is front of mind.

Keywords

Privacy, third-party tracking, library privacy policies, data sharing

In late 2019, Heather Staines and Doreen Bradley became co-editors of a special issue of the journal Serials Librarian that focused on privacy in libraries. The impetus for this special issue emerged from discussions and presentations in privacy at a number of recent professional library conferences. As digital access, network security, shared services, third-party software, and other factors of maturing information systems have transformed the way we work and our users access library resources, privacy concerns have become a much larger and more controversial topic in librarianship. The special issue on privacy was originally mapped out before COVID-19; as the pandemic has moved so many research, teaching, and learning activities online, however, the topic of privacy has really taken on new significance.

Most of the content from the privacy issue of Serials Librarian is already available online. In selecting the content, the co-editors cast a wide net to identify articles that would illustrate the expansive discussion on privacy and libraries. The issue examines third-party software, individualized services, learning analytics, and data sharing. The goal was to feature thought-provoking interviews, articles, and research pieces that highlight new areas of research and new ideas, as well as nuances in current discussions.

There are numerous stakeholders and perspectives that are important to consider in these conversations, including library staff, users, vendors, researchers, and many more. Additionally, we must recognize the continuing issues of racism and equity, and the need to be cognizant not only of the voices that are being heard but equally those that are not yet being heard in these conversations.

Physical-equivalent privacy

We are delighted that three of authors from the special privacy issue could be with us. They will be presenting three very different perspectives around privacy and libraries. The first presenter is Dorothea Salo, who is a distinguished faculty associate at the University of Wisconsin Madison's iSchool. She has an

MS in Spanish and an MLS, both from the University of Wisconsin. She will discuss privacy audits and physical-equivalent privacy.

> Aside from profession-wide ethics codes, libraries have two formal ways of assessing and declaring commitments to privacy. One is via actual privacy policies, such as the excellent one from the San Francisco Public Library and the occasional special statement about a specific issue, such as American Library Association Council's Resolution on the Misuse of Behavioral Data Surveillance in Libraries.[1] These policies and statements indicate what librarians claim they care about with respect to privacy. These statements do not by themselves measure what library policies and practices actually do to live up to the claims they contain.

> Library privacy audits, such as those outlined by ALA, check libraries' claims to privacy protection, measuring how well they actually are protecting privacy.[2] They do have important limitations, however:

> The audit is solely of systems and services completely under the library's control. There is no auditing of non-library third parties who also handle library or patron data, such as campus or municipal IT or an e-resource vendor.

> Audits do not often measure practices against claims in library privacy policies or professional ethics codes. They only present a snapshot of current library practice, without evaluating its appropriateness or conformance to professional ethics.

> Duke Libraries published a detailed and worthwhile privacy audit in 2020 that illustrates these weaknesses in ALA's privacy audit guidelines, detailing the Libraries' privacy practices without evaluating them.[3] Lacking a clear measurement modality,

how could Duke have done otherwise? In the context of ubiquitous, deeply troubling, nearly undetectable new modes of privacy violation, mostly but not entirely online, the lack of a usable yardstick reduces the usefulness of privacy audits significantly. Worse still, it is rare for librarians to know enough about present-day privacy threats – which are often fairly to highly technical, as well as commonly obfuscated – to assess library practices against them.

A thought experiment from a NASIG 2015 Vision Session may point to a useful yardstick.[4] The presentation asked attendees to compare common online behavioral surveillance to a librarian following individual patrons around with a video camera, recording everything they looked at wherever they went. The privacy affordances of an e-resource may be intelligibly measured by determining, to the extent possible, what data is being collected, where the data travel, how identifiable patron behavior is based on it, and then imagining the equivalent collection and dissemination of the same kind of data for a patron using a bound volume in the stacks. This process has been dubbed "physical-equivalent privacy."

There are additional good reasons to do this comparison, beyond the utility of having a yardstick. One is captured in Article I of the ALA Code of Ethics, which states in part, "We provide the highest level of service to all library users through equitable service policies and equitable access." Privacy-violating service is hard to construe as the highest level of service. Article VI reads, "We do not advance private interests at the expense of library users."[5] Few librarians are under any illusion that most systematic patron surveillance advances anything but private interests.

Additional thoughts on applying the physical-equivalent privacy method are available in the article establishing it, as are real-world examples and discussion related to the in-process SeamlessAccess initiative.[6]

Third-party tracking

Next, Gabriel Gardner will discuss third-party tracking. Gardner is an associate librarian and the discovery coordinator at the University Library, California State University Long Beach. He holds a BA in Spanish from Elmhurst College and an MLS from Indiana University Bloomington.

Many roads lead to Rome, and moral and practical problems in librarianship are often soluble via different routes. By happenstance on my part, but through careful foresight of the editors, my contribution to the special issue of Serials Librarian on privacy provides an empirical snapshot of how public libraries are doing on something akin to Dorothea Salo's "physical-equivalent privacy" yardstick. I am no ethical deontologist, judging acts to be good or bad solely upon a set of rules without examining outcomes, yet the eight articles of the American Library Association's Code of Ethics are absent of consequentialist quibbling (except Article IV about so-called intellectual property).[7] They lend themselves to straightforward interpretations without exceptions, and when we look at Articles III, protecting privacy and confidentiality, and VI, not advancing private interests, with deontological eyes, a clear problem with current practices appears.

The commercialized internet is functionally equivalent to a panopticon, or constant surveillance, unless users take proactive measures to preserve some modicum of privacy. Because so much patron-library interaction takes place mediated through

computers using the internet, that means librarians must be hyper-vigilant if we are to even begin to approach physical-equivalent privacy online. Specifically, my paper in the special issue focuses on three closely related problems: third-party tracking, secure connections (i.e., default usage of the HTTPS protocol), and easy discoverability of privacy policies. For details on each of these, readers should consult the full paper; what follows below is a discussion of third-party tracking and why much of it may violate the ALA Code of Ethics.[8]

Third-party tracking is when some entity, other than the website directly visited by the user, tracks the visit to the site. Tracking may simply be a log that the page was loaded but may also involve detailed behavioral data about how a user behaves on a page (e.g., heat maps of clicks or mouse activity), how long she was there, which pages she subsequently navigated to, etc. Such tracking was recently the subject of an ALA resolution and is subsumed under the label of third-party tracking.[9] Methods of tracking go back to 1996 and have been growing in magnitude and sophistication ever since.

How does third-party tracking work? There are several methods, the most common of which is web browser cookies. Those are small files left on your computer in the local storage of the web browser. As of 2016, with enforcement beginning in 2018, cookies have been regulated in the European Union as a result of the General Data Protection Regulation (GDPR). This has resulted in pop-ups that you are probably now accustomed to seeing. But those engaged in third-party tracking have a target-rich environment for browser identification beyond just cookies. The various methods include: Adobe Flash and Microsoft Silverlight local shared object storage (thankfully both are deprecated),

HTTP Strict Transport Security (HSTS) tags, HTTP Entity (ETag) tags, web browser fingerprinting, web beacons (also called tracking pixels), and, of course, your IP address if it is not shared or dynamic.[10] Interested readers can consult the full paper for further details about third-party tracking.[11]

As noted above, there is no exception to ALA Code of Ethics Article VI. Any third-party code, in particular widgets or images from social media companies, operating on your library website has the potential to violate Article VI. In the case of social media companies, we know that those entities track everything. If your catalog sends user queries via the URL then any widgets or images loaded from social media company servers are providing those companies with what users are looking for. This is not to say that libraries cannot participate in social media, only that they must be very cautious in how social media content appears on their websites. All images or content must be hosted locally on library servers, not linked from the source; running social media company code on your library website is an unambiguous violation of Article VI.

Now, to put a little flesh on the theoretical bones assembled above, I'll review how public libraries in the United States and Canada are doing. Using two different browser plugins which measure third-party tracking, Ghostery and Disconnect, I looked at all members in the Urban Libraries Council and the Canadian Urban Libraries Council in 2017. Both tracking detection plugins have different taxonomies of third-party tracking with Disconnect typically recording more tracking but with Ghostery having a more granular classification system. What is the scale of tracking on library websites and catalogs? Using the Disconnect count, I found that 88% of the libraries in the sample are enabling some kind of third-party tracking (the Ghostery figure was 87%). This

seems like an immediate cause for concern until we drill down to note that most of the tracking is the usage of Google Analytics and Google Tag Manager. Some libraries are using widgets that are connected to commercial data brokers (e.g., AddThis), but the most common type of third-party tracking after web analytics came from Facebook code that libraries have inserted onto their sites.

The mere usage of web analytics is not necessarily cause for concern, but it could be if libraries are not informing their users that entities other than the library are recording their usage of and behavior on library websites. This brings me to another aspect of my analysis where I looked for the presence of a link to a privacy policy or terms of use policy on the library homepages. I did not conduct a qualitative analysis of whether the policies were accurate, the question was simply, "Is a link there?" This is the most bare-bones criteria I could come up with. A majority of libraries (58%) lacked an easily discoverable privacy policy from their homepage. That figure combined with the widespread usage of Google Analytics reveals a massive chasm between common practice and ALA Code of Ethics Article III, which states: "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."[12] Ultimately, 49% of libraries in the sample enabled some type of third-party tracking without easily informing their users that it was taking place. The hypothetical ideal situation of a discoverable privacy policy on a website with zero third-party tracking only applied to 3% of the sample.

There is a tremendous opportunity for education on this issue of not just our patrons but also librarians. Simply put, some libraries, particularly those allowing social media

code on their websites, are violating Code of Ethics Articles III and VI. Many more

should work to make their privacy policies discoverable.

Licensing privacy

Finally, we will hear from Lisa Janicke Hinchliffe who will discuss licensed resources and privacy. Janicke Hinchliffe is professor and coordinator of information literacy services and instruction at the University Library, University of Illinois at Urbana Champaign, where she's also affiliate faculty in the School of Information Sciences, and in the Center for Global Studies. She has an MA in educational psychology and an MLS from the University of Illinois.

> *Libraries are increasingly concerned about the ways in which users of library licensed*
>
> *resources are being tracked by the third-party providers of these resources. Efforts such*
>
> *as the "NISO Consensus Principles on Users' Digital Privacy in Library, Publisher, and*
>
> *Software-Provider Systems" laid a foundation for library and provider agreements on*
>
> *user control, data security and management, and transparency.[13] Nonetheless,*
>
> *researchers at Massachusetts Institute of Technology (MIT), the University of*
>
> *Minnesota, and Temple University have documented the extent of the data collection*
>
> *and capture that is happening on these third-party platforms.[14] There are concerns as*
>
> *well that usage data is being aggregated, mapped to other data sets, and possibly even*
>
> *being fed into the commercial data sector, being sold to advertisers and the like.*
>
> *After just a bit of exploration, many librarians find themselves agreeing with Cody*
>
> *Hanson, director of IT at the University of Minnesota Libraries, who has published*
>
> *extensively on privacy and claims, and who noted that it is not possible for libraries to*
>
> *provide, in any meaningful way, assurance of anonymity or privacy for those who are*

using library licensed resources with the current state of user tracking.[15] For librarians, this is deeply troubling given our long-standing commitment to user privacy and confidentiality as embodied in this statement from the ALA Code of Ethics: "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."[16]

The Licensing Privacy project (https://publish.illinois.edu/licensingprivacy/), funded by the Mellon Foundation, seeks to use the power of library licensing agreements to effect change in third-party platform practices in order to bring them into alignment with library values of privacy, confidentiality, and respect for user control over their own data. It reflects an identified Pathway for Action from the IMLS-supported National Web Privacy Forum.[17] The goal of Licensing Privacy is to develop model license language on user privacy that would support libraries in advocating for user privacy when contracting for services and content. By ensuring that user privacy is contractually protected in licensing agreements, service contracts, etc., libraries would be able to hold platforms accountable for their data practices.[18]18 In addition to model language, the project seeks to develop a toolkit of resources that libraries can use in support of the language, contract negotiations, policy analysis, etc.

Staff for the Licensing Privacy project include Lisa Janicke Hinchliffe as Principal Investigator, as well as consultants Danielle Cooper (Ithaka S + R) and Becky Yoose (LDH Consulting Services). The advisory group consists of Ann Okerson (senior advisor on electronic strategies, Center for Research Libraries), Katie Zimmerman (director of copyright strategy, MIT Libraries), and Scott W.H. Young (user experience &

assessment librarian, Montana State University). Additional consultants will be hired as the project progresses.

Originally conceptualized pre-pandemic, the design of the project was necessarily re-developed to reflect the need to work virtually. Instead of an in-person convening, the project is now structured as producing a series of white papers/reports and companion webinars. Planned topics include: library practices in negotiating for privacy; assessing content providers contract compliance with library privacy guidelines, standards, and best practices; and a librarian's primer to authentication and authorization. Dissemination will continue through conference presentations and workshops as travel regulations permit.

As a whole, the Licensing Privacy project will signal the importance libraries place on privacy to content and services providers, enable greater protection of user privacy, and help libraries communicate clearly and accurately to users about the privacy protections they do (and do not) have when using library-licensed resources and services.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Doreen Bradley is Director of Learning Programs and Initiatives at the University Library at the University of Michigan.

Dorothea Salo is a Distinguished Faculty Associate at the iSchool at the University of Wisconsin Madison.

Gabriel J. Gardner is Associate Librarian and the Discovery Coordinator at the University Library, California State University Long Beach.

Lisa Janicke Hinchliffe is Professor and Coordinator of Information Literacy Services and Instruction at the University Library at the University of Illinois at Urbana Champaign. She is also an Affiliate Faculty at the School of Information Sciences and in the Center for Global Studies.

Heather Ruland Staines is Director of Community Engagement and Senior Consultant at Delta Think.

Notes

[1] San Francisco Public Library, "Privacy Policy," August 16, 2018, https://sfpl.org/about/privacy-policy (accessed May 1, 2021); ALA Council, "Resolution on the Misuse of Behavioral Data Surveillance in Libraries," (February 25, 2021), https://www.ala.org/advocacy/intfreedom/datasurveillanceresolution (accessed May 1, 2021).

[2] American Library Association, "Developing or Revising a Library Privacy Policy," (April 25, 2014), https://www.ala.org/advocacy/privacy/toolkit/policy (accessed May 1, 2021).

[3] Joyce Chapman and Angela Zoss, "Duke Libraries Data Privacy and Retention Audit Report," (January 2020), https://dukespace.lib.duke.edu/dspace/handle/10161/20061 (accessed May 1, 2021).

[4] Dorothea Salo and Stephen Kharfen, "Ain't Nobody's Business If I Do (Read Serials)," The Serials Librarian 70, no. 1–4 (2016): 55–61, https://doi.org/10.1080/0361526X.2016.1141629 (accessed May 1, 2021).

[5] ALA Council, "Code of Ethics of the American Library Association," (2008), https://www.ala.org/tools/ethics (accessed May 1, 2021).

[6] Dorothea Salo, "Physical-Equivalent Privacy," The Serials Librarian (2021, in press), https://doi.org/10.1080/0361526X.2021.1875962.

[7] ALA Council, "Code of Ethics of the American Library Association."

[8] Ibid.

[9] ALA Council, "Resolution on the Misuse of Behavioral Data Surveillance in Libraries."

[10] "HTTP cookie," n.d., in Wikipedia, https://en.wikipedia.org/wiki/HTTP_cookie (accessed May 1, 2021).

[11] Gabriel Gardner, "Aiding and Abetting: Third-Party Tracking and (In)secure Connections in Public Libraries," The Serials Librarian (2021, in press), https://doi.org/10.1080/0361526X.2021.1943105.

[12] ALA Council, "Code of Ethics of the American Library Association."

[13] NISO, "NISO Consensus Principles on Users' Digital Privacy in Library, Publisher, and Software-Provider Systems," (2015), http://www.niso.org/publications/privacy-principles (accessed May 1, 2021).

[14] Micah Altman and Katie Zimmerman, "Privacy Gaps in Mediated Library Services," CNI Fall Meeting, Washington, DC (2018), https://www.cni.org/topics/user-services/evaluating-and-closing-privacy-gaps-for-online-library-services (accessed May 1, 2021); David Lacey and Cody Hanson, "Collecting, Correlating, Stitching, Enriching: How Commercial Publishers are Creating Value by Profiling Users," CNI Spring Meeting (2019), https://www.cni.org/topics/identity-management/collecting-correlating-stitching-enriching-how-commercial-publishers-are-creating-value-by-profiling-users (accessed May 1, 2021).

[15] Cody Hanson, "User Tracking on Academic Publisher Platforms," (2019), https://www.codyh.com/writing/tracking.html (accessed May 1, 2021).

[16] ALA Council, "Code of Ethics of the American Library Association."

[17] Scott W. H. Young, Sara Mannheimer, Jason A. Clark, and Lisa Janicke Hinchliffe, A Roadmap for Achieving Privacy in the Age of Analytics: A White Paper from A National Forum on Web Privacy and Web Analytics (2019), https://doi.org/10.15788/20190416.15445 (accessed May 1, 2021).

[18] Lisa Janicke Hinchliffe and Katie Zimmerman, "Negotiating for What We Want: A Proposal for Model License Language on User Privacy," CNI Fall Meeting, Washington, DC, 2018, https://www.cni.org/topics/user-services/evaluating-and-closing-privacy-gaps-for-online-library-services (accessed May 1, 2021).