2022

The Privacy Librarian is In! How Privacy Issues Affect Researchers and Libraries

John Felts
Coastal Carolina University, jfelts@coastal.edu

Heather Staines
Delta Think, heather@seamlessaccess.org

Tim Lloyd
LibLynx, tim@liblynx.com

Keondra Bailey
NISO, kbailey@niso.org

Wilhelmina Randtke
Georgia Southern University, wrandtke@georgiasouthern.edu

Recommended Citation

Felts, John; Staines, Heather; Lloyd, Tim; Bailey, Keondra; and Randtke, Wilhelmina, "The Privacy Librarian is In! How Privacy Issues Affect Researchers and Libraries" (2022). *Library Faculty Publications*. 13.
https://digitalcommons.coastal.edu/lib-fac-pub/13

The Privacy Librarian is In! How Privacy Issues Affect Researchers and Libraries

John Felts, Head of Information Technology & Systems, Coastal Carolina University, jfelts@coastal.edu, https://orcid.org/0000-0002-7104-9878

Heather Staines, Senior Strategy Consultant, Delta Think, Heather.Staines@deltathink.com, https://orcid.org/0000-0003-3876-1182

Tim Lloyd, CEO, LibLynx, tim@liblynx.com, https://orcid.org/0000-0003-0495-5691

Keondra Bailey, Assistant Standards Program Manager, NISO, keonbai@gmail.com

Wilhelmina Randtke, Head of Systems and Technologies, Georgia Southern University, wrandtke@georgiasouthern.edu, https://orcid.org/0000-0002-7439-8205

Abstract

Faced with an increasingly complex online environment through which libraries provide access to scholarly resources, librarians have found it difficult to educate users in protecting their personal information and online behaviors from inappropriate and sometimes unauthorized use while promoting the personalization services that users find beneficial.

Modeled after the long-running Peanuts cartoon with Lucy offering advice for 5 cents, a panel composed of librarians, a vendor, and a publisher convened an interactive session that tackled key privacy issues in the researcher, vendor, and library framework. It began with the "Privacy Librarian" training a new library employee while a stream of patrons asked them privacy-related questions and research scenarios.

After this tour through privacy issues, the panel highlighted the context in which these patron questions existed. Topics included privacy considerations in complex authentication schemas, website tracking, browser security, campus surveillance, and data privacy in third-party vendor accounts.

Complex Authentication

The current complexity of authentication makes it particularly challenging for librarians to manage data privacy. The first challenge of many worth highlighting is that different authentication methods share different levels of personal data as a result of vendors varying in their support for authentication methods due to their industry focus and their technical architecture. For example, vendors that primarily serve the academic community are more likely to support IP address authentication while vendors that primarily support the corporate sector are more likely to support authentication via username and password. As a result, few if any libraries rely on a single authentication method. When a patron asks what data is shared when they access library resources, the answer typically depends on which resources are being used and how they are accessed. Examples of the varying levels of data shared include personal IP addresses that are used in IP authentication and referring URLs, personally-identifying email addresses used in username/password authentication, and personal data such as a name, email address, and employment information that are sometimes shared in single sign-on.

Secondly, some methods of single sign-on can be configured to share different levels of personal data with different library vendors. While this is considered beneficial because it allows librarians to selectively share personal data only when necessary, it also causes the data sharing landscape to be more complex. For example, federated authentication can be configured to release no personal data at all for access to some resources while allowing a name and email address for other resources that rely on email communication.

Another challenge is that traditional contract language dates back to a simpler world of IP authentication of usernames and passwords and typically doesn’t address the more complex and nuanced data sharing practices of more modern technologies like federated authentication. Because contracts themselves can often languish unchanged for long periods of time, these deficiencies can continue long after a library has adopted more modern solutions. One industry initiative helping in this area is the SeamlessAccess Contract Language working group, which has developed new template language for inclusion in the Center for Research Libraries Library Model License Agreement.[1]

Website Tracking

Universities are increasingly using website tracking tools to track their student’s website behaviors. Every activity is recorded, logged, and reviewed in a practice called learning analytics, and data analysts are using this information to predict whether a student could potentially struggle with their courses or eventually even drop out of school altogether.

Advocates state that analyzing student’s website behaviors and other analytics can help to provide a more personal learning experience, and can mitigate drop-out rates by looking at how usage patterns of online learning resources can be an early indicator of academic performance. Others warn against a massive security breach or for the data to be used in such a way that users find inappropriate. There’s a general consensus that more transparency is required about

[1] https://seamlessaccess.org/posts/2022-10-31-contractlanguagemodellicense/

what data is being collected and about the potential inferences that can be made by those who collect this data.[2]

When a library user asks about the ramifications that Google tracking code snippets are found on all university web pages, librarians typically find that they can only respond that the library isn't the primary selector and administrator of the content management systems and analytics software utilized on campus since these decisions come by authority of university IT. Therefore, librarians can only continue to develop relationships with the university IT department to hopefully create opportunities to advocate for user privacy while remaining cognizant of the issues around website tracking. A key resource for gaining a better understanding of this issue is Marshall Breeding's research in Library Technology Reports (v. 55, no. 7) which explores the issues and technologies needed to deploy a library website with adequate protections for the privacy of those who visit.[3]

Browser Security

When a library user reports that they are unable to access content on an Apple device that they were able to successfully access on a different computer, librarians are finding that changes in browser and browser engine technologies that are being implemented by big tech to bolster privacy are having the unintended consequence of disrupting authentication.

To give an example of this behavior and how to potentially troubleshoot it, Duke University Press (DUP) rolled out a new program called the Scholarly Publishing Collective, which is a collaborative agreement between Michigan State University Press, Penn State University Press, Society of Biblical Literature Press, and University of Illinois Press to host selected journals from each publisher. In the months leading up to the program's rollout, DUP ingested thousands of customer's data involving organizations both new and known. Most new customers were eager to utilize this access, so they would contact DUP directly to review their holdings information and set up their access and administrative contact information.

One of these institutions was a small, professional institution located in the UK. They had a subscription to one e-journal and wanted to use a referring URL so that their students could access the content via Moodle (an open source management system), which is where things got tricky. When testing the connection from a wired, campus-based desktop the connection was seamless. However, when connecting via mobile phone network with a personal laptop, the connection would not work and access was denied. They could still view the article on the website, but access was denied upon the redirect opening a new tab and trying to download a PDF of the article from the resulting page.

Troubleshooting this type of e-resource issue is heavily dependent on what browser settings the representative has in place and asking for access information such as details about IP or Shibboleth/OpenAthens isn't enough. In this case, the test occurred in Google Chrome. Chrome blocks third-party cookies by default in incognito mode, but users also have the option to block them in a regular session. When clicking a link to the PDF in DUP's humanities platform, the

[2] https://www.theguardian.com/higher-education-network/2016/aug/03/learning-analytics-universities-data-track-students

[3] https://journals.ala.org/index.php/ltr/issue/view/738

user is redirected to a new window to download the content. However, when this redirect occurs the proxy connection breaks and the access token is lost. Big tech is just starting to look at this issue and to identify potential solutions.

Campus Surveillance

The topic of surveillance in schools and universities is one that has garnered controversy in recent years. While these are implemented under the guise of improving student scholarship and overall well-being, surveillance tools such as CCTV cameras and security systems with facial recognition software, metal detectors, building access controlled by RFID-enabled swipe cards, and tracking online user behaviors can also be seen as an invasion of student privacy and an undermining of their civil liberties. Surveilling students can act as a crime deterrent and also serves administrative interests such as in the awarding of scholarships, reputation of academic programs, and generating positive statistics for the school. But there is concern that the ever-increasing intrusiveness of these sophisticated tracking systems will infantilize students in the very place they're expected to grow into adults, and that universities are trading off future worries for the immediacy of convenience, comfort, and ease.[4]

When a student enquires about security cameras in their library, librarians find that they are routinely left out of these conversations and have little to no knowledge of what technologies are being used and how information about student behaviors is stored and utilized. Librarians have long-advocated for confidentiality when using library resources which seems to be in direct conflict with increasing university surveillance of its students. However, seldom do librarians have the ability to affect policy changes at the highest levels of university administration. But at least in understanding how, when, and what types of surveillance technologies are being utilized on campus, librarians can stand ready to inform students, highlight awareness, and where appropriate advocate for change.

Data Privacy

When a library user creates accounts on numerous vendor platforms when using IP access, the library has no ability to intervene on the user's behalf. They login to a resource via their proxy server that's likely integrated with an LDAP or Azure directory, then routinely create a personal user account on any number of vendor platforms. The library or the university do not have the ability to protect a user's privacy or intervene in any way should user privacy be violated.

However, with federated access user identity is baked into the authentication and authorization process when first logging into the vendor's platform. This pertains to the entity categories for use in the configuration of federated authentication systems that outline which attributes about the user are passed from a subscribing organization to the service provider. In addition, two new categories were approved by REFEDS early this year which gives libraries and vendors the

[4] https://www.washingtonpost.com/technology/2019/12/24/colleges-are-turning-students-phones-into-surveillance-machines-tracking-locations-hundreds-thousands/

technical specifications needed to manage the sharing of user information and protect user privacy:[5]

- The Anonymous Authorization Entity Category passes no personally identifiable information about a user, only authorizes access to a resource and nothing more

- The Pseudonymous Authorization Entity Category also grants access but also passes a pseudonymous identifier that obfuscates user identity but still allows personalization services that benefit the user

And because technical solutions are many times not enough, SeamlessAccess convened a Contract Language Working Group to help libraries address the lack of useful contract language found in many if not most vendor contracts, because so many of these pre-date modern technologies and have little useful information about browser behavior and single sign-on technologies.[6]

[5] https://refeds.org/a/2558

[6] https://seamlessaccess.org/posts/2022-03-14-contractlangmodellicense/