Data from Library Implementations

In order to assess the current state of practice in the way that libraries handle patron privacy, observations were made for a selection of libraries. The two groups selected for the study included the members of the Association of Research Libraries and the largest 25 public libraries in the United States. The selection of these two groups focuses the study toward the largest and most sophisticated libraries. These libraries are more likely to have the technical capacity and the financial resources to implement products that meet a high level of functional requirements. Smaller libraries may have fewer resources to configure or adjust their technology products relative to privacy concerns. This exercise hypothesized that these groups of libraries would exhibit the most sophistication in their websites and catalogs, both in terms of features and in attention to privacy and security.

Methodology

The study relies on lists of libraries belonging to the two groups of interest. The members of the Association of Research Libraries are listed on the organization's website, and a list can be generated from the libraries.org resource on Library Technology Guides.

Association of Research Libraries
www.arl.org

Libraries.org resource
http://librarytechnology.org/libraries/arl
or
http://librarytechnology.org/libraries./search.pl?ARL=on

The list of the largest 25 public libraries in the United States was based on the ALA Fact Sheet, "The Nation's Largest Public Libraries: Top 25 Rankings."[1] An expanded version of the ranking table, provided in table 3.1, includes the integrated library system and online catalog product implemented by each organization.

Each of the websites in the two lists was visited in the last week of December 2014, noting several characteristics:

- Is the website itself delivered using HTTP or HTTPS?
- What is the primary discovery interface or online catalog presented? Many of the ARL libraries feature both a discovery interface and a traditional online catalog. Some have multiple discovery interfaces, though the one presented as the default search tool is the one considered. These public libraries generally do not have article-level discovery services, so only the online catalog was considered.
- Does the discovery interface use HTTPS by default?
- Does the online catalog use HTTPS by default?
- Using the Ghostery plug-in for Chrome, all tracking mechanisms detected on the website, online catalog, or discovery interface were noted. The following notation can be seen on tables 3.2 and 3.3. (Abridged forms of the tables appear here; the full tables are available online.)
  a = all major components, including website, catalog, and discovery interface
  d = discovery interface only
  c = online catalog only
  w = website only

Ghostery plug-in for Chrome
https://www.ghostery.com

A variety of tracking mechanisms were noted. Most, if not all, send some data to an external organization. Whether that data includes personally identifiable information would require additional technical analysis and tracing. At a minimum, these tracking mechanisms report externally that a specific resource or page associated with the specific web server was accessed at a specific time.

- Google Analytics
- Google Ajax search API
- Google AdSense
- Google Translate
- Google Tag Manager
- DoubleClick (owned by Google)
- Yahoo Analytics
- Adobe Omniture Analytics
- Adobe Tag Manager
- Adobe TypeKit
- Facebook Connect
- Facebook Social Plugin
- Twitter Button
- AdThis
- Piwik Analytics
- Crazy Egg
- WebTrends
- New Relic

Observations

Data collection for this study was performed in November and December 2015, with all data reviewed and revised in the last week of December. The data collected is meant to represent only a snapshot reflecting current practices at that specific time. It is expected that many of the sites may change even by the time this report is published. This data can also serve as a baseline to measure any changes that might take place in the future. Any such changes would serve as a barometer of whether the concerns related to patron privacy increase or diminish, at least as measured by the implementation of secure resource delivery and through the use of tagging mechanisms related to external commercial entities.

Large Academic Libraries

- Out of 124 ARL member libraries considered, only 16 (13%) present their main website using HTTPS.
- Out of the 95 ARL member libraries that feature an online catalog search on their website, only 12 (14%) default to HTTPS for search activity.

Table 3.1. The 25 largest US public libraries, included in this study

Los Angeles Public Library, CA
New York Public Library
County of Los Angeles Public Library, CA
Chicago Public Library, IL
Brooklyn Public Library, NY
Queens Borough Public Library, NY
Miami-Dade Public Library System, FL
Houston Public Library, TX
Harris County Public Library, TX
Broward County Libraries Division, FL
San Antonio Public Library, TX
Orange County Public Libraries, CA
Free Library of Philadelphia, PA
Phoenix Public Library, AZ
Las Vegas-Clark County Library District, NV
Hawaii State Public Library System, HI
King County Library System, WA
Sacramento Public Library, CA
San Diego Public Library, CA
Hillsborough County Public Library Cooperative, FL
Dallas Public Library, TX
San Bernardino County Library, CA
Riverside County Library System, CA
Hennepin County Library, MN
Orange County Library District, FL

- Out of the 100 ARL member libraries that feature a discovery service on their website, only 17 (17%) default to HTTPS for search activity.
- Out of 124 ARL member libraries considered
  - All 124 included some form of tracking tag to an external commercial entity.
  - 119 include Google Analytics page tagging on their main website.
  - 11 include Google AdSense advertising tracking tags on their main website or discovery interface.
  - 22 include DoubleClick advertising tracking tags on their main website or discovery interface.
  - 37 include New Relic tags on their discovery interface. ProQuest Summon consistently embeds New Relic.

Large Public Libraries

- Out of the 25 large public libraries considered, only 2 (8%) present their main website using HTTPS.
- Out of the 25 large public libraries considered, only 7 (28%) use HTTPS by default for catalog search activity.
- Of these 7 secure catalogs, 5 base their catalog search on BiblioCore.
- 24 out of the 25 (96%) embed Google Analytics tags for their website and catalog.

Table 3.2. This table shows data from the largest public libraries on the security of their catalog and website, as well as whether Google Analytics is in place. This is an abridged form of the larger table, reviewing the use of other analytics tools, available from the Library Technology Guides website. http://librarytechnology.org/web/breeding/ltr-52-4-table3-2

	Website	Catalog	Secure?	Google Analytics
Los Angeles Public Library, CA	n	LS2 PAC	n	wc
New York Public Library	n	Encore	n	wc
County of Los Angeles Public Library, CA	n	eLibrary	n	wc
Chicago Public Library, IL	n	BiblioCommons	y	wc
Brooklyn Public Library, NY	n	BiblioCommons	y	wc
Queens Borough Public Library, NY	n	Local	n	wc
Miami-Dade Public Library System, FL	n	PowerPAC	n	wc
Houston Public Library, TX	n	Portfolio	n	wc
Harris County Public Library, TX	n	Portfolio	n	wc
Broward County Libraries Division, FL	n	LS2 Pac	n	wc
San Antonio Public Library, TX	n	WebPac Pro	n	wc
Orange County Public Libraries, CA	n	Enterprise	y	wc
Free Library of Philadelphia, PA	n	VuFind	y	wc
Phoenix Public Library, AZ	y	PowerPAC	n	wc
Las Vegas-Clark County Library District, NV	n	WebPac Pro	n	wc
Hawaii State Public Library System, HI	n	Enterprise	n	wc
King County Library System, WA	n	BiblioCommons	y	wc
Sacramento Public Library, CA	y	Encore	n	wc
San Diego Public Library, CA	n	BiblioCommons	y	wc
Hillsborough County Public Library Cooperative, FL	n	PowerPAC	n	wc
Dallas Public Library, TX	n	PowerPAC	n	wc
San Bernardino County Library, CA	n	PowerPAC	n	w
Riverside County Library System, CA	n	Powerpac	n	wc
Hennepin County Library, MN	n	Local?	y	wc
Orange County Library District, FL	n	WebPac Pro	n	

w = website
c = catalog

- 4 out of the 25 embed DoubleClick advertising tracking tag.
- 1 out of the 25 embeds Google AdSense advertising tracking tag.
- 7 embed Facebook Connect.

The results of this study are nothing short of alarming relative to the privacy practices seen in these elite groups of institutions. Despite the findings in chapter 2 that all of the systems available have the technical capacity to be deployed using encrypted secure communications, only small percentages of these libraries have implemented it for their online catalogs or discovery services. Almost as few implement their websites with security, which is also standard capability of commercial and open-source web servers or content management systems. These sites are also promiscuous in their use of commercial tracking agents. Almost all use Google Analytics. Only one site, the University of Albany, was observed with no detectible tracking agents. The use of commercial advertising tracking agents from Google AdSense and DoubleClick is also noteworthy.

It should also be noted that the major commercial services and social networks employ HTTPS, including Facebook, Twitter, and all Google services.

The lack of pervasive implementation of secure communications use cannot be blamed on the lack of capability in the systems, but rather may be attributed only to gaps in awareness of its benefits or lack of expertise to reconfigure existing implementations. Vendors and libraries could partner to reshape the security landscape quickly if this were identified as a priority.

The public exposure of network traffic can be considered as only one small component of an overall strategy in the way that technology systems used in a library environment protect patron privacy. How technology infrastructure handles patron data and search behaviors can be seen as the foundation needed to support higher-level features and services that may also have privacy implications. Libraries may, for example, want to provide social features that enable their patrons to opt into sharing information about themselves and their reading habits, either with selected groups of other patrons or publicly. Libraries might want to collect additional

Table 3.3. This table, running multiple pages in its full form, shows security findings from each ARL library's website, catalog, and discovery service, along with use of Google Analytics. The full data set showing all analytics tools found can be downloaded from the Library Technology Guides website. http://librarytechnology.org/web/breeding/ltr-52-4-table3-3/

	Website	Catalog	Secure?	Discovery Interface	Discovery Secure?	Google Analytics
Arizona State University	y	WebPac Pro	n	Summon	y	a
Auburn University Libraries	n	VuFind	n	none		a
Boston College	n			Primo	n	a
Boston University	n			Primo	n	a
Boston Public Library	n			BiblioCommons	y	a
Brigham Young University	n	eLibrary	n	Local	y	a
Brown University	n	Blacklight	y	Summon	y	a
Case Western Reserve University	n	WebPac Pro	n	Summon	n	a
Center for Research Libraries	n	WebPac Pro	n			a
Colorado State University	n			VuFind	n	a
Columbia University	n			Blacklight		a
Cornell University	y			Blacklight	y	a
Dartmouth College	n	WebPac Pro	n	Summon	n	a
Duke University	n	Aleph	n	Drupal/ Summon		a
Emory University	n			Primo	n	a
Florida State University	y	Mango	n	Summon	n	a
George Washington University	n	Drupal	n	Drupal/ Summon	n	a
Georgetown University	n	WebPac Pro	n	Summon	n	a
Georgia Institute of Technology	n			Primo	n	w
Harvard University	n	Aleph	n	Primo	n	w
Howard University	n	WebVoyage	n	Summon	n	dc
Indiana University	y	Blacklight	n	Drupal EDS API	y	a
Iowa State University	n			Primo	n	a
Johns Hopkins University	n	Blacklight	y	Blacklight	y	a
Kent State University	n	WebPac Pro	n	EDS	n	a
Louisiana State University	n	eLibrary	n	EDS	n	a
Massachusetts Institute of Technology	n	EDS	y	EDS	y	a
McGill University	n	Aleph	n	WorldCat	n	a
McMaster University	n			VuFind	n	a
Michigan State University	n	WebPac Pro	n	Summon	n	a
National Archives and Records Administration	n					w
National Research Council Canada	n	WebPac Pro	n			a
New York State Library	n					a
New York University	n	Primo	n	Xerxes / EDS	y	a
New York Public Library	n	Encore	n			a
North Carolina State University	n			Local	n	a
Northwestern University	n			Primo	n	a
Ohio State University	y	WebPac Pro	n	Worldcat	n	a
Oklahoma State University	n	Primo	n	Summon	n	a
Pennsylvania State University	y			Summon	y	a
Princeton University	n			Blacklight/ Primo	n	a
Purdue University	y			Primo	n	a
Queen's University	n	WebVoyage	y	Summon	n	a
Rice University	n	eLibrary	n	Drupal EDS API	n	a

Table 3.3. (cont.)

	Website	Catalog	Secure?	Discovery Interface	Discovery Secure?	Google Analytics
Rutgers University	n	VuFind	y	EDS		a
Smithsonian Institution	n	iPac	n	Summon	n	a
Southern Illinois University	n	VuFind	y	EDS	?	a
Stony Brook University	n	Aleph	n	EDS	n	a
Syracuse University	n	WebVoyage	n	Summon	n	a
Temple University	n	WebPac Pro	y	Summon	n	a
Texas A&M University	n	WebVoyage	n	EDS		a
Texas Tech University	n	Primo	n	EDS	n	a
Tulane University	n	WebVoyage	n	Primo	n	a
Library of Congress	n	Local	n			
National Agricultural Library	n	Voyager	n			a
National Library of Medicine	y	Voyager	n	PubMed	n	
Universite Laval	n			Ariane	n	a
University at Albany	n	Aleph	n	EDS		
University at Buffalo	n	VuFind	n	Summon	n	a
University of Alabama	n	WebVoyage	n	Drupal EDS API	n	a
University of Alberta	n	eLibrary	n	EDS	n	a
University of Arizona	n	WebPac Pro	n	Summon	n	a
University of British Columbia	n	WebVoyage	n	Summon	n	a
University of Calgary	n			Drupal / Summon	n	a
University of California -- Berkeley	n	WebPac Pro	n	EDS	n	a
University of California -- Davis	y	Aleph	y			a
University of California -- Irvine	n	WebPac Pro	n			a
University of California -- Los Angeles	n	WebVoyage	n	Summon		a
University of California -- Riverside	n	WebPac Pro	n	EDS	n	a
University of California -- San Diego	n	WebPac Pro	n			a
University of California -- Santa Barbara	n	Aleph	n			a
University of Chicago	n	VuFind	y	EDS API		a
University of Cincinnati	n	WebPac Pro	n	Summon	n	a
University of Colorado -- Boulder	n	Encore	n			a
University of Connecticut	n			Primo	n	a
University of Delaware	n			WorldCat	n	a
University of Florida	n	Mango	n	Summon	n	a
University of Georgia	n	VuFind	n	EDS	n	a
University of Guelph	n			Primo	n	a
University of Hawaii -- Manoa	n			Primo	n	a
University of Houston	n	WebPac Pro	n	Primo	n	a
University of Illinois -- Chicago	n	VuFind	n	Summon	n	w
University of Illinois at Urbana-Champaign	n	VuFind	n	Local?	n	a
University of Iowa	n	Aleph	n	Primo	n	a
University of Kansas	y	Voyager	n	Primo	n	a
University of Kentucky	n	WebVoyage	n			a
University of Louisville	n			WorldCat	y	a
University of Manitoba	n	Primo	n	Primo	n	a
University of Maryland	n	Aleph	n	WorldCat	y	a
University of Massachusetts -- Amherst	n	Aleph	n	WorldCat	n	a
University of Miami	n	WebPac Pro	n	Summon	n	a
University of Michigan	n	VuFind	n	Drupal	n	a
University of Minnesota -- Twin Cities	y	Primo	n	Primo	n	a
University of Missouri -- Columbia	n	WebPac Pro	n	Summon		a

Table 3.3. (cont.)

	Website	Catalog	Secure?	Discovery Interface	Discovery Secure?	Google Analytics
University of Montreal	n	Primo	n	Primo	n	a
University of Nebraska -- Lincoln	n	WebPac Pro	n	Encore	n	a
University of New Mexico	n			WorldCat	y	a
University of North Carolina -- Chapel Hill	n	Endeca	n	Local	n	a
University of Notre Dame	n	Primo	n	Primo		a
University of Oklahoma	y	Primo	n	Primo	n	w
University of Oregon	n	Primo	n	Primo	n	w
University of Ottawa	n	WebPac Pro	y	Primo	n	a
University of Pennsylvania	n	Local	n	Local	n	a
University of Pittsburgh	n	Voyager	n	Summon	n	a
University of Rochester	n	WebVoyage	n	Summon	n	dc
University of Saskatchewan	n	WebPac Pro	n	Primo	n	a
University of South Carolina	n	WebPac Pro	n	Encore		a
University of Southern California	y			Summon	n	
University of Tennessee -- Knoxville	y			Primo	n	a
University of Texas -- Austin Libraries	n	WebPac Pro	n	Summon	n	a
University of Toronto	y	Local?	n	Summon API	y	a
University of Utah	n			Primo	n	a
University of Virginia	n	eLibrary	n	Blacklight	n	
University of Washington	n			Primo	n	a
University of Waterloo	n	Primo	n	Local	n	a
University of Western Ontario	n	WebPac Pro	n	Summon	n	a
University of Wisconsin -- Madison	n	Local?	y	Primo	n	a
Vanderbilt University	n	Primo	n	Local / Primo	n	a
Virginia Tech	n	WebPac Pro	n	Summon	n	a
Washington State University	n	Primo	y	Primo	y	a
Washington University in St. Louis	n	WebPac Pro	n	Metalib / Primo	n	a
Wayne State University	y	WebPac Pro	n	Local / Summon	y	a
Yale University	n	WebVoyage	n	Local	n	a
York University	n	eLibrary	n	VuFind	y	a

non-anonymized data to create value-added services. Yet, without a secure foundation, it may be difficult to manage such services without exposing more private data than intended.

This report reveals a very uneven reality as seen in library websites, catalogs, and discovery environments related to secure transmission, a baseline requirement for patron privacy, and in the leakage of data regarding visits to library resources to commercial entities. Repeating annually the data collection described in this chapter would provide an interesting measure of whether the library community concurs with the concerns raised and is able to institute the changes needed to secure their resources and contain exposure to tracking agents. The Electronic Frontier Foundation and others are working toward improving privacy on the web at large via increased adoption of HTTPS. Given the emphasis that libraries give privacy in their ethics and policies, it would be reasonable to expect them to be leaders rather than laggards in that trend.

Related Projects and Resources

NISO Consensus Framework to Support Patron Privacy

Funded through a grant from the Andrew W. Mellon Foundation, NISO conducted a participatory process to investigate issues related to the privacy and security of systems employed by libraries and to develop a set of statements addressing key topics to help inform library practices. The project included a series of virtual discussions carried out with invited participants via webinar, a two-day in-person meeting in San Francisco, and a phase of synthesizing the information collected into a report. The project addressed perspectives of systems provided by libraries, vendors (such as integrated library systems and discovery services), and publishers. The final report, including twelve statements of “privacy principles,” titled *NISO Consensus Principles on Users’ Digital Privacy in Library, Publisher, and Software-Provider Systems (NISO Privacy Principles)* was published December 10, 2015, and is available online from NISO.

NISO Consensus Principles on Users’ Digital Privacy
www.niso.org/apps/group_public/download.php/15863/NISO%20Consensus%20Principles%20on%20Users%92%20Digital%20Privacy.pdf

Library Digital Privacy Pledge

The Library Freedom Project has launched an initiative it calls the Library Digital Privacy Pledge, soliciting libraries to commit to the delivery of their web-based resources through HTTPS. The Library Freedom Project was founded and is directed by Alison Macrina with contributions from other volunteers. The efforts of this initiative to champion the need for libraries to encrypt transmission of their web resources is consistent with the topic of this report. The Library Freedom Project received funding from the Knight Foundation News Challenge on Libraries.

Library Digital Privacy Pledge
https://libraryfreedomproject.org/ourwork/digitalprivacypledge

EFF: Let’s Encrypt

The Electronic Frontier Foundation has led an initiative called Let’s Encrypt, aimed at facilitating encryption on the web for all types of sites. The project provides tools to reduce the cost and effort of enabling encryption on a site, such as providing a free service, available since December 2015, to create valid digital certificates.

Let’s Encrypt
https://letsencrypt.org

Note

1. “ALA Library Fact Sheet 13: The Nation’s Largest Public Libraries: Top 25 Rankings,” American Library Association, August 2014, www.ala.org/tools/libfactsheets/alalibraryfactsheet13.

Other Resources

American Library Association. “An Interpretation of the Library Bill of Rights, Privacy.” Accessed January 10, 2016. www.ala.org/advocacy/intfreedom/librarybill/interpretations/privacy.

Blum, Dan. “Privacy by Design and the Online Library Environment.” *Information Standards Quarterly* 26, no. 3 (Fall 2014): 4–11. www.niso.org/publications/isq/2014/v26no3/blum.

Breeding, Marshall. “Perspectives on Patron Privacy and Security.” *Computers in Libraries* 35, no. 5 (June 20015): 12–14. http://librarytechnology.org/repository/item.pl?id=20831.

Mayer, Jonathan R., and John C. Mitchell. “Third-Party Web Tracking: Policy and Technology.” In *Proceedings: 2012 IEEE Symposium on Security and Privacy, S&P 2012,* 413–27. Los Alamitos, CA: IEEE Computer Society, 2012. http://dx.doi.org/10.1109/SP.2012.47. Available online at https://jonathanmayer.org/papers_data/trackingsurvey12.pdf.

Noh, Younghee. “Digital Library User Privacy: Changing Librarian Viewpoints through Education.” *Library Hi Tech*, 32, no. 2 (2014): 300–17. http://dx.doi.org/10.1108/LHT-08-2013-0103.

Sturges, Paul, Vincent Teng, and Ursula Iliffe. “User Privacy in the Digital Library Environment: A Matter of Concern for Information Professionals.” *Library Management*, 22, no. 8/9 (2001), 364–70. http://dx.doi.org/10.1108/01435120110406309.