Newsletter

ALA TechSource
alatechsource.org

Smart Libraries™

Formerly Library Systems Newsletter™

50 East Huron Street, Chicago, Illinois 60611-2795, USA

ISSN 1541-8820 May 2017 Volume XXXVII Number 5

Smarter Libraries through Technology

Protecting Patron Privacy

By Marshall Breeding

The ability for the technology systems of libraries to safeguard the privacy of patrons continues to be an issue of utmost importance. The library profession as a whole upholds the privacy of their patrons as they make use of library-provided resources. Most libraries also have privacy policies stating what data they collect about their patrons and how that data will be used or shared. Those values tend to be well understood and enforced in the realm of physical materials. How libraries safeguard patron privacy within their virtual presence appears to be much more uneven. In this issue of *Smart Libraries Newsletter*, we discuss the complexities of protecting patron privacy in both the physical and virtual arenas and present the results of a mini-study of the current state of practice of patron privacy on the websites of a selected group of large libraries. Unfortunately, this study reveals a fairly lax level of implementation of the technical mechanisms necessary to protect patron privacy as they make use of library resources via their library's websites and catalogs.

Safeguarding Circulation and Patron Profile Data

Protecting the confidentiality of circulation data related to the materials patrons borrow from libraries has become a well-established practice. The basic concern lies in eliminating any data beyond what is operationally necessary that identifies the materials a patron has borrowed. Libraries generally consider this data confidential. They treat any data that reveals the items consulted or borrowed by a patron with great care.

The circulation module of the integrated library system (ILS) manages data related to the inventory of content items owned by the library and maintains detailed records for each of its patrons eligible to borrow materials. Bibliographic records describe the title, subject, and other details, while item or holdings records the location of each copy and the library's unique identifiers. Patron records include significant personal information, such as name, address, and age. Many libraries also record sensitive identification numbers such as a driver's license or social security number. In some cases, credit card and financial data may be stored or managed through a PCI-compliant payment system.

When a patron borrows an item from the library, the circulation module will make a link between the patron record and the content item. The circulation transaction binds the item and patron records so the library knows what items are in use and therefore not available to other borrowers. The library also needs sufficient data to send notices to the patron when an item has not been returned by its due date as well as for other routine communications. Most libraries consider the data linking a borrower with an item as an operational necessity for active loans.

Once an item has been returned, the link between the item and the patron can be dissolved. From the perspective of managing circulation functions, this association is no longer essential. Some libraries, however, may choose to retain some information

IN THIS ISSUE

regarding the immediately previous borrower in case some problem arises, such as discovering that the item has become damaged.

Data relating to a circulation transaction can exist in a variety of locations. The circulation module will create a transaction record including the patron identifier, the item record number, and timestamps for when the item was borrowed or returned. Some systems will write the patron identifier into the item record as a convenient way for staff members to know the current or previous borrower. Patron records may include a list of items currently or previously borrowed. Transaction logs may exist as part of the disaster recovery procedures so that any database actions that occurred subsequent to the most recent backup can be restored. For web-based interfaces, the web server logs may also include entries created when the item was borrowed or returned. In some cases, network infrastructure components may also log transaction. Removing all traces of a circulation transaction may require a series of technical tasks, which in most cases can be automated.

At a minimum, libraries will take measures to ensure that no personally identifiable link is preserved once a patron has returned an item. Libraries usually need to maintain detailed statistics related to items borrowed to produce reports and to analyze borrowing and collection trends. Most ILS circulation modules include functionality to remove personally identifying data from transaction records, which makes the patron data anonymous while preserving category and demographic data to satisfy reporting requirements. This functionality may—or may not—also scrub all related log files to likewise remove personally identifying data. Many libraries also make use of third-party analytics services and utilities that operate on the basis of data exported from the library's ILS or other systems. Libraries need to be aware of whether the data supplied to these analytical tools includes personally identifiable information.

The privacy features of an ILS operate on multiple levels. During the time of an active check-out, the system should enforce authorization features that limit access to the details of a circulation transaction to supervisory personnel and preclude access to others. This level of control prevents a lower-level circulation desk worker from viewing the items a patron has charged or to reveal that information if asked by a third party. Some supervisory or administrative personnel will usually be able to access this information. Once an item has been returned, an ILS configured for a high level of privacy would eliminate the possibility of gaining access to that transaction regardless of administrative authorization or technical access to the system. In the event of a legal request from a third party or an unauthorized system intrusion, such data cannot be provided or accessed since it no longer exists.

Such tight control of circulation data may run contrary to the interest some libraries haves in enabling features to provide personalized or social capabilities. Patrons may expect, for example, to be able sign into their account and see lists of all the materials they have previously borrowed. The ability to provide this information to the patron likewise makes it available to authorized disclosures by the library or unauthorized access in the event of a security event. Libraries can create opt-in policies in which patrons consent to the preservation of their borrowing history and inform them of the associated risks. Libraries have to balance the opposing pressures of providing personalized services, embracing many of the concepts of social networks with the contradictory elements of eliminating any data that may compromise the privacy of their patrons.

Confidentiality of Patron Transactions Conducted Online

Library catalogs, discovery interfaces, and websites can expose sensitive information regarding patron interactions with the library's online services. The communications between the patron and the library's technical systems conveys sensitive data, including the query entered, lists of results, record selected, and items of content viewed online or downloaded. The data describing patron online behavior contains data even more sensitive than circulation transactions. Search queries, for example, could be captured or requested as evidence of interest in a topic even when the patron did not find or select specific content items.

Web-based interfaces transmit all the data related to a patron's interactions with a library catalog or discovery services across local networks or the internet. The URL generated by the search interface will include a query string, encoded in name/value pairs, describing the term entered into the search box, qualifiers, and other details. Interfaces configured to use the GET directive will transmit this data directly as part of the URL; those using POST will send the same information via a temporary file. Regardless of the method used, this data can be captured on the network if it is sent without encryption.

Information transmitted via local networks or the internet can be vulnerable to interception by parties other than the intended recipient. Technical tools are readily available to capture traffic on wired or wireless networks by those with physical access or proximity. Enterprise networks on campuses or for corporations may include diagnostic or monitoring tools capable of capturing transmissions. It is also reasonable to assume that internet service providers have the technical ability to intercept or record all the data that traverses their networks.

Mandate to Encrypt

Given the reality of the ease of interception of data transmitted across networks, encryption stands as the key strategy for ensuring the privacy of that data. End-to-end encryption ensures that only the intended senders and receivers are able to decrypt and view data. Fortunately, the means for enabling encryption are readily available, inexpensive, and can be implemented with only a moderate level of technical ability. For web-based services, the implementation of the HTTPS protocol encrypts the data stream between the originating server and the user's browser. The activation of HTTPS requires that the web server operator obtain a digital certificate from a digital certificate authority that authoritatively identifies the service and provides the public and private security keys needed for encryption. Once a valid digital certificate has been installed on the web server, it can then be configured to communicate using the encrypted HTTPS protocol instead of HTTP, which sends data as clear text without protection. The managers of the service may also need to review its content to ensure that any links or included content are coded to use HTTPS instead of HTTP. Pages with a mixture of HTTP and HTTPS content will be considered as insecure by most web browsers.

Libraries need to ensure that the web servers associated with their website, catalogs, discovery interfaces, or content resources preserve the privacy of their patrons as they use the resources the libraries provide. Each of the services in the library's web presence may have a different set of procedures or complications in the activation of HTTPS, often involving tasks by both the library's technical staff and the vendor providing the software or its hosting services. In most cases, the library will need to acquire the digital certificates involved. How those certificates are installed and the technical configuration of the services to use HTTPS will vary.

The online catalog module of an ILS or a discovery interface can be more complex to configure to use HTTPS than a basic website. The developers of these products will need to ensure that all aspects of functionality operate properly and consistently deliver secured web pages when HTTPS is activated. The scenario may also differ for local implementations versus those hosted by the vendor. Some discovery interfaces are deployed using the domain name of the vendor, along with the responsibility of obtaining digital certificates. Some vendors may implement HTTPS by default; for others, it may be an optional configuration that must be requested by the library.

From a privacy perspective, it is also important for HTTPS to be mandatorily enforced. It is possible for a service to be configured to use HTTPS by default, but to also function with HTTP. Any links to the service coded with HTTP or if the user happens to type HTTP would transpire without protection. Sites that automatically redirect requests made with HTTP to HTTPS offer more comprehensive privacy protection.

The internet increasingly has become more aggressive regarding collecting information that intrudes on the privacy of individuals. In the consumer arena, there are implicit or explicit understandings that the provision of free services comes at the cost of personal information, which may then be used commercially, such as for targeted advertising. Advertising networks have deep roots in the infrastructure of the web, gathering details of sites visited, items selected, or any other clues that may point toward consumer interests. At an even deeper level, the possibility of internet service providers recording and selling their customer's web browsing history, even without their permission, has deep implications for privacy. For organizations such as libraries that desire to protect the confidentiality of their patrons who make use of their services, mandatory encryption is essential.

The Current State of Privacy Practice in Large Library Organizations

Securing the actions performed on a web-based service via HTTPS has become well established and may soon become the expected norm. Organizations dealing with financial data, medical records, or other confidential information routinely implement security via HTTPS. Social networks such as Facebook and Twitter, all the services offered by Google, and most news sites now operate entirely over HTTPS. Consistent with this growing expectation for security and privacy, many web browsers, such as Chrome, now display information or warning indicators for any site not using HTTPS.

In the context of the prevailing expectation that reputable websites are deployed using HTTPS, libraries lag behind other types of organizations in a wholesale shift toward providing this level of privacy and security for their web-based services.

As a follow-up to the May/June 2016 issue of *Library Technology Reports* on "Privacy and Security for Library Systems," which included data gathered in the last quarter of 2015, we have surveyed the websites of several libraries noting whether selected services use HTTP or HTTPS. The sites reviewed included the members of the Association of Research Libraries (ARL) and a selection of large public libraries. These organizations would more likely have the technical and financial resources to manage their web-based services according to current expectations for privacy and security. As some of the largest libraries in North America, they are also most likely to have adopted privacy policies and the technical expertise to implement them.

For each of the organizations selected, we tabulated the status of the library's main website, the primary online catalog, and discovery service. These services remain within the control of the library and can be considered key indicators of the level of privacy possible as patrons make use of the library's online services. Especially for academic libraries, some may offer a link to the online catalog of their ILS as well as an index-based discovery service. The concept of the online catalog does not apply to those using products such as Alma or WorldShare Management Services. The discovery interface provides access to both local holdings and article-level content. There are a small number of academic libraries among those reviewed that offer an online catalog but not a discovery service.

Our review of the web-based resources of these libraries (shown in Tables 1 and 2 on the following page) reveals improvements beyond what was observed in 2015, but shows that the majority continues not to enforce encryption to protect patron privacy. Out of the 124 ARL member libraries considered, only 42 percent present their website using HTTPS. Out of the 25 major public libraries considered, only 36 percent deploy their website using HTTPS. Table 3 breaks down Table 2, revealing the 25 major public libraries in the United States and looking at the security of each library's catalog and website. The following two charts on page 6 shows the percentage of ARL member libraries employing each of the different catalog and discovery services available.

In the context of the prevailing expectation that reputable websites are deployed using HTTPS, libraries lag behind other types of organizations in a wholesale shift toward providing this level of privacy and security for their web-based services. This observation is surprising given the concern libraries state regarding patron confidentiality and privacy. As libraries work to improve their technical infrastructure, those that value the privacy of their patron's use of the online services will want to give high priority to the implementation of HTTPS for the systems under their control.

Beyond the values of the library profession to protect patron privacy, the urgency of this change is also driven by upcoming changes in the way that browsers flag page security. Google's Chrome browser already displays an informational message for sites presented through HTTPS: "your connection

Table 1: ARL Member Libraries

	2015			2017		
	Website	Catalog	Discovery	Website	Catalog	Discovery
Total	124	95	100	124	107	107
HTTPS	16	12	17	52	31	26
Percent HTTPS	13%	13%	17%	42%	29%	24%

Table 2: 25 Major Public Libraries in the United States

	2015		2017	
	Website	Catalog	Website	Catalog
Total	25	25	25	25
HTTPS	2	7	9	12
Percent HTTPS	8%	28%	36%	48%

Table 3: Major Public Libraries in the United States

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

Figure 1: Catalog Service ARL Member Libraries are Employing

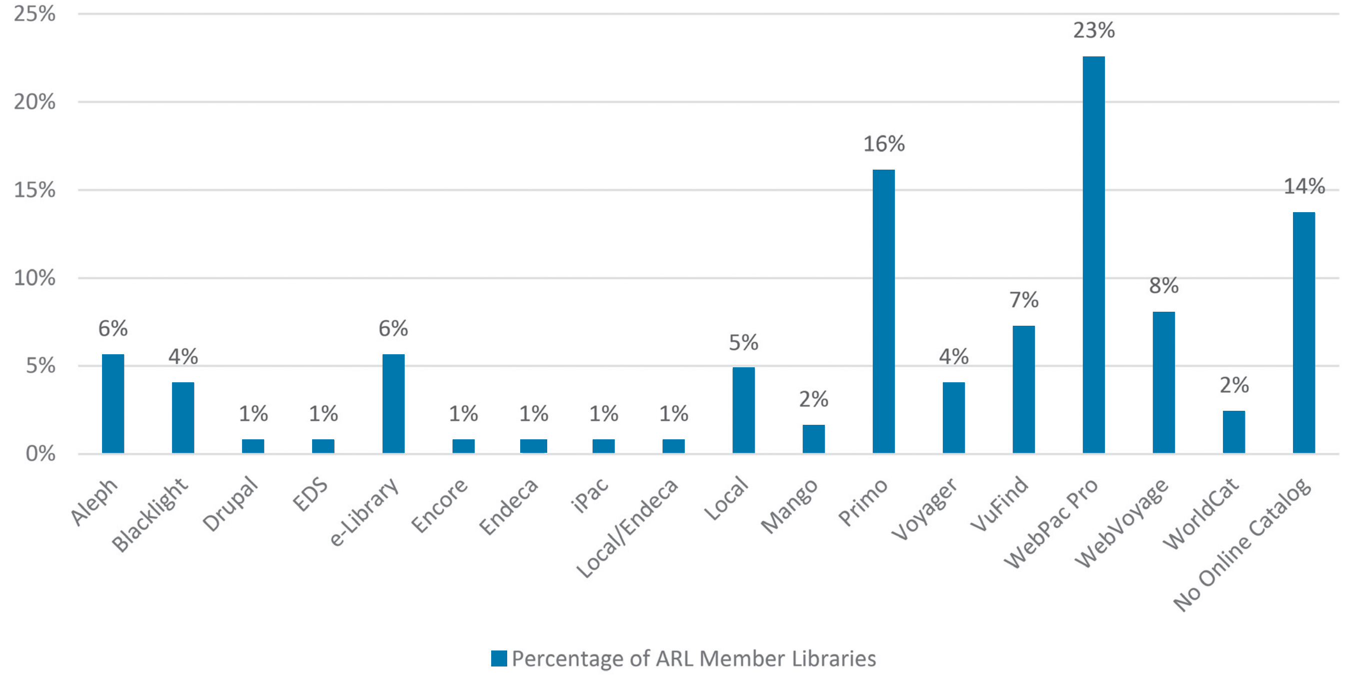
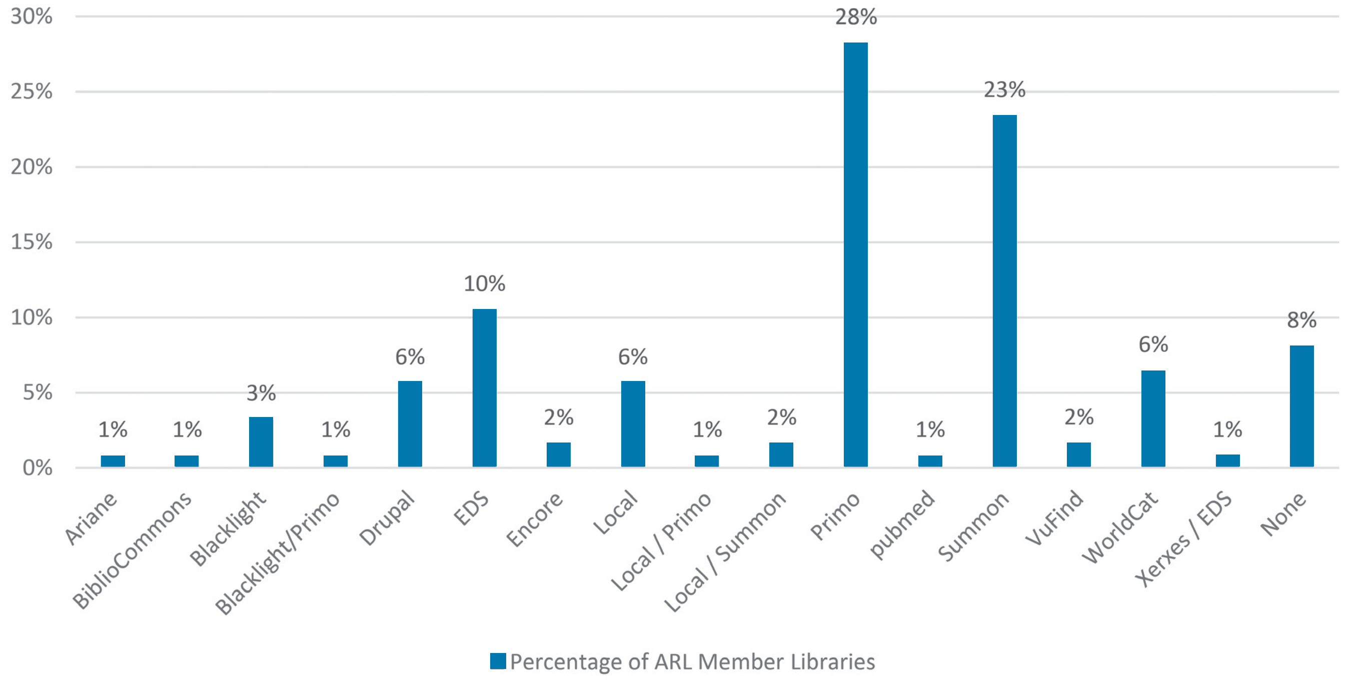


Figure 2: Discovery Services ARL Member Libraries are Employing



to this site is not secure." Although no specific date has been set, Google states that future versions of its browser will elevate the warning with a conspicuous red "Not secure" indicator. I would urge libraries to move rapidly toward comprehensive use of HTTPS for their web-based resources in advance of this change if they want their resources to be perceived as trusted and reliable.

Smart Libraries Q&A

Each issue, Marshall Breeding responds to questions submitted by readers. Have a question that you want answered? Email it to Samantha Imburgia, Associate Editor for ALA TechSource, at simburgia@ala.org.

Are open source systems inherently more secure than proprietary library automation products? What are some of the issues that libraries need to consider when choosing systems related to security?

It is my experience that both open source and proprietary systems can be implemented at very high levels of security. Likewise, without careful attention to technical implementation issues, any type of software can become vulnerable.

Open source takes a philosophically different approach to security than proprietary software. Since anyone can view the underlying coding of an open source product, it may be possible for hostile programmers to scrutinize it for vulnerabilities that can be exploited. Conversely, the entire development and user community of an open source application can likewise continually inspect the source code to detect and repair vulnerabilities. Naturally, the development community of an open source project has to be diligent, attune to security issues, and able to create and deploy security patches rapidly. Proprietary software places the responsibility for security on the organization responsible for its development. Since the source code is not available for public inspection, there may be times where vulnerabilities may exist but remain undetected. But when a security issue is discovered, the systems developers must act rapidly to create and deploy a fix to resolve the vulnerability before it can be exploited. In either case, the worst-case scenario would be a "zero day" vulnerability, where a vulnerability exists, is known by malicious agents, and no fix has been developed to protect the systems involved. In the event of a zero-day attack, developers must work very rapidly to create and deploy a fix and to mitigate any exposure or damage to data that may have taken place during the interval when the vulnerability was being exploited.

Among organizations such as libraries that may lack adequate technical support, there may be many implementations of systems based on older versions of the software in which all of the security patches available may not have been applied. Regardless of whether they use open source or proprietary software, libraries should strive to keep all their systems updated with the latest versions of the software. This approach will not only offer more protection from security vulnerabilities, it will mean that all recently-developed features or fixes to functionality will be available. Libraries are especially lax in implementing new versions of software in order to minimize disruption, deferring such updates to periods of slower activity. A regular practice of implementing minor updates as they become available should result in more stable and secure systems to support the library's operations and services.

ALA TechSource
alatechsource.org

Smart Libraries Newsletter
American Library Association
50 East Huron Street
Chicago, IL 60611-2795 USA
Address Service Requested

May 2017
Smarter Libraries through Technology

Smart Libraries Newsletter

Marshall Breeding's expert coverage of the library automation industry.

Editor
Marshall Breeding
marshall.breeding@librarytechnology.org
Twitter: @mbreeding

Managing Editor
Samantha Imburgia
312-280-3244
simburgia@ala.org

Digital Access for Subscribers
journals.ala.org/sln

Production and design by the American Library Association Production Technology Unit.

Smart Libraries Newsletter is published monthly by ALA TechSource, a publishing imprint of the American Library Association.

alatechsource.org